DETECTION OF POLYMORPHISM IN XENOBIOTIC METABOLISING ENZYMES USING DATABASE ANALYSIS

Natasha Luan Tetlow, B.Sc. (Hons)

A thesis submitted for the degree of Doctor of Philosophy
at the
Australian National University

Division of Molecular Medicine
John Curtin School of Medical Research
Australian National Univeristy
Canberra, Australia

January, 2003

DECLARATION

I declare that the work presented in this thesis, unless otherwise acknowledged, is the original work of the author and has not been presented for a degree elsewhere.

Natasha Tetlow

January, 2003

ACKNOWLEDGMENTS

In 1998, I started a PhD project in the now defunct Cancer Genetics Group led by Dr Maija Kohonen-Corish. Unfortunately, this project was destined to be shortlived, with Maija leaving JCSMR for Sydney eight months later. Still determined to complete a PhD degree, I approached Professor Philip Board about potential projects in the Molecular Genetics Group. I would like to thank him for accepting me into the group and for his continual support throughout the last four years, especially during the last year when I decided to get married and run away to London just as I was starting to write this thesis. It has been a long year, but I appreciate the patience and effort that has been required to get comments back re the thesis via email and fax.

I would like to thank Dr Anna Robinson, who joined the Molecular Genetics Group towards the end of my project, for her invaluable advice, enthusiasm and help with various protein related questions. Thank you so much for helping me with the protein modelling - I have learnt a lot in the past few months! Whilst on the subject of proteins, I would also like to thank Dr Tim Mantle, who was a visitor in the group for some months, for his advice regarding setting up the GSTM3-3 kinetics experiments. I could never have done these without your expert advice and enthusiasm on the subject of proteins. Finally, I would like to thank Astrid Whitbread, a fellow PhD student, for carrying out the GST Omega recombinant protein work and enzymatic characterisations on the GSTO1-1 variants.

Thanks also to the rest of the Molecular Genetics group. Marj Coggan in particular has always somehow found time to provide advice and help with all manners of questions. I would also like to thank Susan Tan for providing the Chinese and African samples used in this study, and Kavitha Kugathas for her assistance in genotyping the African Bantu population.

A Huge thanks must go to Yvonne, Cindy, Alice and Astrid. You have all been fantastic friends ☺. I miss all the coffee breaks, lunches, dinners, trips to the coast, etc - all necessary distractions from a PhD! Also, thanks to Matt Taylor – it was a pleasure sharing our lab with you.

And to my husband Tony who has been willing to sacrifice his weekends and evenings whilst I completed this thesis – THANK YOU THANK YOU THANK YOU. I know it has been tough, but I am eternally grateful for your patience, support, thesis related advice, and also for letting me get away with so much ☺.

Finally, I would like to thank my parents and Marnie, who have always supported my decisions and shown an interest in my work. Thank you for always being there for me.

ABSTRACT

The link between genomic variation and disease was first recognised over 50 years ago, fuelling research into the underlying genetic basis of rare and complex genetic disorders, disease susceptibility and altered drug response. Interest in polymorphism studies has undergone a rapid revival with the recent completion of the Human Genome Project, the advent of sequence databases and the development of computer-based tools designed to detect polymorphisms, all of which are expected to facilitate this area of research.

Of particular interest in the study of disease susceptibility and adverse drug responses are enzymes involved in the metabolism of xenobiotics such as the glutathione *S*-transferases (GSTs). The GST superfamily belongs to the group of phase II detoxification enzymes that facilitate the detoxification of various carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress with which humans are daily faced. Eight classes of soluble human GSTs, with as many as five isoforms per class, have been defined (Alpha, Kappa, Mu, Omega, Pi, Sigma, Theta and Zeta). Although there is some overlap between classes, each GST isoform possesses unique substrate specificities hence the GSTs are able to provide an effective defence against the potentially damaging effects of a range of xenobiotics and endogenously produced toxic compounds.

Polymorphisms in the GSTs have the potential to alter an individual's susceptibility to carcinogens and toxins, and influence the efficacy and toxicity of drug treatment. A number of polymorphisms have been described in the human GSTs and whilst some appear innocuous, others have been implicated in altering susceptibility to diseases such as cancer and Parkinson's Disease, and influencing response and resistance to drugs, simply through altered activity and expression levels. As these polymorphisms have generally been identified by chance, the extent of genetic polymorphism in the human GST genes remains unknown. The recent developments in bioinformatic approaches to polymorphism detection were therefore applied to the GST superfamily. The following chapters describe the discovery of polymorphisms in the Alpha (Chapter 3), Mu (Chapter 4) and Omega (Chapter 5) class GSTs using this approach.

Five polymorphisms were identified in the Alpha class GSTs – the silent *GSTA1* p.K125K polymorphism; p.P110S, p.T112S and p.E210A in *GSTA2* and p.L71I in *GSTA3* (Chapter 3). Although the *GSTA2* polymorphisms had previously been recognised as sequence variants, they had not been the subjects of further investigations. Allele frequencies for each of the five variants were determined in four populations (Australian European, Bantu Africans, Creole Africans and Southern Chinese) representing three of the world's major ethnic groups – Caucasians, Chinese and Africans. Of these, the *GSTA2* p.T112S and *GSTA3* p.L71I polymorphisms were found to be quite rare. Enzymatic characterisation revealed significant alterations in enzyme function were introduced by the *GSTA2* p.S110 variant.

Two Mu class polymorphisms in the *GSTM3* gene – p.G147W and p.V224I, were identified and characterised in Chapter 4. Again, despite previous recognition, these polymorphisms have not been extensively characterised. Genotyping results for these two polymorphisms, in addition to a known intronic deletion, revealed a distinct difference between the allelic profiles of the Southern Chinese compared to that of the Australian European and two African populations. The p.V224I polymorphism was found to be relatively common, with the p.V224 variant being the common variant in the Australian European and African populations, whereas the p.I224 variant was common in the Southern Chinese population. The known intronic deletion was also found to be relatively common in the Australian European and African populations, yet was absent from the Southern Chinese population. In contrast, the extremely rare p.G147W polymorphism was only identified in three Southern Chinese individuals. Enzymatic characterisation revealed that the p.I224 variant was responsible for significant functional changes, but the p.W147 variant did not dramatically alter enzyme function. However, when the p.W147 and p.I224 polymorphisms were combined, a significant decrease in specific activity was observed.

Three novel polymorphisms in the newly characterised GST Omega class genes are described in Chapter 5 – p.A140D and p.E155del in *GSTO1* and p.N142D in *GSTO2*. Whilst the p.A140D *GSTO1* and p.N142D *GSTO2* were relatively common in the four populations genotyped, the p.E155del *GSTO1* polymorphism was quite rare, and only one individual in the Southern Chinese population was found to be homozygous for this deletion. Unfortunately due to protein purification problems, further investigation into

the *GSTO2* polymorphism was prevented. Through enzymatic characterisation of the *GSTO1* allelic isoforms, it was demonstrated that the p.E155del polymorphism was responsible for a significant increase in transferase and thiolreductase activities. In contrast, the p.A140D polymorphism did not appear to have any functional effects on GSTO1-1, however as a substrate unique to GSTO1-1 has not yet been identified, and its monomethylarsonous reductase activity was unable to be measured, the possibility that this polymorphism may alter enzyme activity must be kept in mind.

This study has revealed a moderate level of genetic polymorphism in the GST Alpha, Mu and Omega genes through utilisation of a bioinformatic approach, an approach that has proved to be rapid and relatively efficient. Characterisation of the enzymatic properties has shown that the structure and function of the common allelic isoforms of *GSTA2*, *GSTM3* and *GSTO1* are significantly altered by four of the polymorphisms described here, each of which may exert significant clinical implications.

ABBREVIATIONS

4HNE	4-hydroxynon-2-enal
13-cRA	13-*cis*-retinoic acid
Δ^5-AD	Δ^5-androsten-3,17-dione
ABI	Applied Biosystems
AFB_1	aflatoxin B_1-8,9-*exo*-epoxide
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
APS	ammonium persulfate
ARMS	amplification refactory mutation system
ASK	apoptosis signal-regulating kinase I
BAC	bacterial artificial chromosome
BCNU	1,3-bis(2-chloroethyl)-1 nitrosourea
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BPDE	7β,8 α-dihydroxy-9 α,10α-oxy-7,8,9,10-tetra hydrobenzo(a)pyrene
BSA	bovine serum albumin
CDC	Center for Disease Control
CDE	1,2-dihydroxy-3,4-oxy-1,2,3,4-tetra hydrochrysene
cDNA	complementary DNA
CDNB	1-chloro-2,4-dinitrobenzene
CEPH	Centre d'Etudes Polymorphism Humain
CGAP	Cancer Genome Anatomy Project
CGAPC	CGAP candidate list
CGAPCo	CGAP confirmed list
CGAPV	CGAP validated list
CHIP	Children's Hospital Informatics Program
CLIC	chloride intracellular channel
CRC	colorectal cancer

cSNP	coding region single nucleotide polymorphism
CuOOH	cumene hydroperoxide
DBE	dibromoethane
dbEST	the Expressed Sequence Tag Database
dbSNP	the Single Nucleotide Polymorphism Database
DCA	dichloroacetate
DCM	dichloromethane
DCNB	1,2-dichloro-nitrobenzene
DDBJ	DNA Data Bank of Japan
DDCT	D-dopachrome tautomerase gene
ddH_2O	distilled deionized water
DGGE	denaturing gradient gel electrophoresis
DHLP	dihydrolipoic acid
DHPLC	denaturing high performance liquid chromatography
DMA^V	dimethyl arsinate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithioerythritol
EA	ethacrynic acid
EBI	European Bioinformatics Institute
EDTA	ethylenediamine-N,N,N',N'-tetraacetic acid
EGP	Environmental Genome Project
EMBL	European Molecular Biology Laboratory
EMD	Enzymatic Mismatch Detection Assay
EPNP	1,2-epoxy-3-(*p*-nitrophenoxy)propane
EST	expressed sequence tag
eSTS	expressed sequence tagged sites
FAA	fumarylacetoacetate
GAI	Genetic Annotation Initiative
GDB	GDB™ Human Genome Database

GI	gastrointestinal
GPx	glutathione peroxidase
GPx-I	selenium dependent glutathione peroxidase
GPx-II	selenium independent glutathione peroxidase
GSH	glutathione
G-site	glutathione binding site
GSSG	oxidised glutathione
GST	glutathione *S*-transferase
GSTA	glutathione *S*-transferase Alpha
GSTK	glutathione *S*-transferase Kappa
GSTM	glutathione *S*-transferase Mu
*GSTM1*0*	*GSTM1* null allele
GSTO	glutathione *S*-transferase Omega
GSTP	glutathione *S*-transferase Pi
GSTS	glutathione *S*-transferase Sigma
GSTT	glutathione *S*-transferase Theta
*GSTT1*0*	*GSTT1* null allele
GSTZ	glutathione *S*-transferase Zeta
H-bond	hydrogen bond
HCA	heterocylic amines
HEDS	2-hydroxyethyl disulfide
HGBASE	Human Genome Bi-Allelic Sequences
HGVbase	Human Genome Variation database
HGVS	Human Genome Variation Society
H_2O_2	hydrogen peroxide
H-site	hydrophobic substrate binding site
HUGO	Human Genome Organisation
I.M.A.G.E.	Integrated Molecular Analysis of Genomes and their Expression
IPTG	isopropylthio-β-D-galactosidase
JNK	Jun N-terminal kinase

Kb	kilo base
LB	Luria broth
LLNL	Lawrence Livermore National Library
MA	maleylacetone
MAA	maleylacetoacetate
MAPEG	Membrane Associated Protein in Eicosanoid and Glutathione metabolism
MMA^{III}	monomethylarsonous acid
MMA^{V}	monomethylarsonic acid
mRNA	messenger ribonucleic acid
MPP	multiple presentation phenotype
MRP	multidrug-resistance-associated protein
NADPH	nicotinamide adenine dinucleotide phosphate
NaPi	sodium phosphate buffer
NBD-Cl	7-chloro-4-nitrobenzo-2-oxa-1,3-diaxole
NCBI	National Center for Biotechnology Information
N.D.	not detected
NHGRI	National Human Genome Research Institute
NIEHS	National Institute of Environmental Health Sciences
N.S.	not studied
OD	optical density
OMIM	Online Mendelian Inheritance in Man
PACs	P1-derived artificial chromosomes
PAGE	polyacrylamide gel electrophoresis
PAH	polycyclic aromatic hydrocarbon
PCR	polymerase chain reaction
PD	Parkinson's Disease
PDB	protein databank
PDR	Polymorphism Discovery Resource

PG	prostaglandin
PGDS	prostaglandin D synthase
PhIP	2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine
*p*NBC	*p*-nitrobenzyl chloride
*p*NPA	*p*-nitrophenylacetate
*p*NPB	*p*-nitrophenethyl bromide
PPARγ	peroxisome proliferator-activated receptor γ
RFLP	restriction fragment length polymorphism
ROS	reactive oxygen species
RyR	ryanodine receptor
RyR1	ryanodine receptor 1
RyR2	ryanodine receptor 2
SAHC	*S*-adenosyl-L-homocysteine
SAM	*S*-adenosyl-L-methionine
S.D.	standard deviation
SDS	sodium dodecyl sulfate
S.E.	standard error
SNP	single nucleotide polymorphism
SSCP	single strand conformation polymorphism
SSE	sum of squares
STACK	Sequence Tag Alignment and Consensus Knowledge
TAE	tris/acetate/EDTA/buffer
TBE	tris/borate/EDTA/buffer
TE	tris/EDTA
TIGR	The Institute for Genomic Research
T_m	(melting) annealing temperature
t-RA	*trans*-retinoic acid
TSC	The SNP Consortium
UCSC	University of California at Santa Cruz
UDP	uridine diphosphate

URL	uniform resource locator
UTR	untranslated region
UUGC	University of Utah Genome Center
UV	ultra violet
UWGC	University of Washington Genome Center
WHO	World Health Organisation
YY1	yin yang 1

PUBLICATIONS

CONFERENCE ABSTRACTS

Tetlow N.L., Blackburn A.C., Board P.G. (1999). Novel glutathione transferase polymorphisms identified by EST database analysis. Australian Society for Biochemistry and Molecular Biology. Queensland, Australia. Poster presentation.

Tetlow N.L., Blackburn A.C., Board P.G. (1999). Novel alpha class glutathione transferase polymorphisms identified by EST database analysis. Canberra Hospital Annual Scientific Meeting. Canberra, Australia. Poster presentation.

Tetlow N.L., Blackburn A.C., Chelvanayagam G., Board, P.G. (2000). Alpha class glutathione transferase polymorphisms identified by EST database analysis. Canberra Hospital Annual Scientific Meeting. Canberra, Australia. Poster presentation.

Tetlow N.L., Blackburn A.C., Chelvanayagam G., Board, P.G. (2001). Alpha class glutathione transferase polymorphisms identified by EST database analysis. *Chemico-Biological Interactions*. **133**: 123-5.

Tetlow N.L. and Board P.G. Identification and characterisation of polymorphic variants of mu class glutathione transferases. (2001). Australian Society for Biochemistry and Molecular Biology. Canberra, Australia. Poster presentation.

PAPERS

Board P., Tetlow N., Blackburn A., Chelvanayagam G. (2000). Database Analysis and Gene Discovery in Pharmacogenetics. *Clin Chem Lab Med.* **38(9)**: 863-7.

Board P.G., Chelvanayagam G., Jermiin L.S., Tetlow N., Tzeng H.F., Anders M.W., Blackburn A.C. (2001). Identification of novel glutathione transferases and polymorphic variants by expressed sequence tag database analysis. *Drug Metab Dispos.* **29(4 Pt 2)**: 544-7.

Tetlow N., Liu D., Board P. (2001). Polymorphism of human alpha class glutathione transferases. *Pharmacogenetics.* **11(7):** 609-17.

Whitbread AK, Tetlow N, Eyre HJ, Sutherland GR, Board PG. Characterization of the human Omega class glutathione transferase genes and associated polymorphisms. *Pharmacogenetics.* In Press.

TABLE OF CONTENTS

CHAPTER 2 – MATERIALS AND METHODS

CHAPTER 3 – THE ALPHA CLASS GSTS

Chapter 1

Literature Review

1.1 Introduction

The goal of this thesis was to identify novel polymorphisms in xenobiotic metabolising enzymes using database analysis. In order to provide an appropriate overview of this topic, the literature review is divided into two sections. The first section will review the application of database analysis to the area of polymorphism detection, and the second will provide a general review of the glutathione transferases, the family of xenobiotic metabolising enzymes that were the subject of the detailed investigation presented in this thesis.

1.2 Polymorphism Detection Using Database Analysis

1.2.1 Introduction

With the exception of monozygotic twins, every individual is genetically unique. Individual response to the daily barrage of drugs, toxins and other environmental or dietary compounds with which humans are constantly faced, and the resultant effects, can differ dramatically due to this genetic variation. The link between genomic

"The link between genomic variation and disease was recognised over 50 years ago [Ingram, 1957], and it has since been recognised that some variations have physiological consequences in disease susceptibility and adverse drug reactions. However, a resurgent interest in detection of these variations, in particular single nucleotide polymorphisms (SNPs), was not fuelled until the recent sequencing of the human genome, the advent of sequence databases and the capacity to detect SNPs *in silico*."

genome with heritable phenotypes and are likely to facilitate large-scale association studies. SNP data is estimated to have a major impact on population genetics, drug development, forensics, cancer and genetic disease research [Brookes, 1999].

1.2.2 WHY STUDY POLYMORPHISMS?

Sequence variations – inserts, deletions and single base substitutions – in genomic DNA are collectively known as polymorphisms. A locus is considered polymorphic if at least two alleles exist at that particular locus and the allele frequency in a population of the least common variant is >1%, although this guideline is not always strictly followed [Brookes, 1999]. Single base substitutions, also known as SNPs, account for approximately 90% of DNA polymorphism [Collins *et al.*, 1998] and current estimates suggest there is a potential ten million SNPs in the human genome [Kruglyak & Nickerson, 2001; Lai, 2001]. Although they have been calculated as occurring every 1000 bp [Cooper *et al.*, 1985; Kwok *et al.*, 1996], the number of nucleotide differences between two chromosomes, known as the nucleotide diversity [Nei & Li, 1979], varies between chromosomal regions and also between ethnic groups [Horton *et al.*, 1998; Nachman *et al.*, 1998; Nickerson *et al.*, 1998; Kruglyak & Nickerson, 2001; The International SNP Map Working Group, 2001]. Hence some regions are relatively devoid of polymorphism and others, such as the HLA class I and MHC II regions, are polymorphic hotspots [Guillaudeux *et al.*, 1998; Horton *et al.*, 1998].

The majority of polymorphisms are found within non-coding regions [Cooper *et al.*, 1985; Nickerson *et al.*, 1998] and generally do not affect protein function. However, changes close to intron-exon boundaries may affect splicing of mRNA, resulting in the deletion of one or more exons, or the creation of cryptic splice sites that result in incomplete or inactive alternative transcripts [Krawczak *et al.*, 1992]. Some promoter region polymorphisms may affect protein function indirectly through altering the rate of transcription, altering the function of regulatory sequences that control gene expression or the stability or processing of the mRNA transcript of the gene.

Polymorphisms that influence protein function are more likely to occur in the coding sequence of a gene, and each gene is estimated to have four to six of these coding region polymorphisms [Collins *et al.*, 1998; Cargill *et al.*, 1999]. Approximately 40%-50% of the SNPs found in coding regions (cSNPs) are non-synonymous – missense mutations that change amino acid residues [Cargill *et al.*, 1999; Halushka *et al.*, 1999]. Approximately 60% non-synonymous cSNPs have a minor allele frequency below 5%

and are more likely to be restricted to specific ethnic groups [Cargill *et al.*, 1999]. It is believed that the low proportion of these cSNPs possibly reflects selection against deleterious alleles during evolution [Cargill *et al.*, 1999; Halushka *et al.*, 1999]. Only an estimated 33%-36% of these non-synonymous polymorphisms are non-conservative [Li & Sadler, 1991; Cargill *et al.*, 1999] and it is these that are most likely to either affect protein function or stability, depending on the location within the protein, or in the case of premature stop codons, shorten the protein. Recent studies have found that SNPs in residues involved in maintaining protein stability are significantly more predominant than those directly involved in binding and catalysis, and are more likely to be associated with disease and other deleterious effects through overpacking, loss of H-bonds, backbone strain and buried charge residues [Wang & Moult, 2001; Ramensky *et al.*, 2002]. The effects polymorphisms have on a protein can therefore range from subtle effects that are only observed under certain circumstances, to changes that have a significant phenotypic effect.

1.2.3 Polymorphism Detection Methods

1.2.3.1 Laboratory-Based Techniques

In 1979 the first SNP in a human gene underlying a genetic disorder was reported, a nonsense mutation in the β-globin gene responsible for causing β-thalassemia [Chang & Kan, 1979]. This prompted further research into the genetic basis of disease, and a number of phenotype directed techniques were developed to discover SNPs that contributed to the aetiology of these diseases. These methods are effective for identifying highly penetrant single alleles responsible for rare, single locus diseases such as cystic fibrosis [Kerem *et al.*, 1989; Rommens *et al.*, 1989] and Huntington's Disease [Gusella *et al.*, 1983; The Huntington's Disease Collaborative Research Group, 1993]. All of these procedures are based on the amplification of the target DNA sequence followed by detection of sequence variations using electrophoresis based methods that compare either alterations in sequence conformation such as the Single Strand Conformation Polymorphism (SSCP) [Orita *et al.*, 1989], Denaturing Gradient Gel Electrophoresis (DGGE) [Sheffield *et al.*, 1989] and Heteroduplex Analysis [Keen

et al., 1991] methods; or the ability of the product to be cut with different enzymes or chemicals such as the Restriction Fragment Length Polymorphism (RFLP), the Enzymatic Mismatch Detection Assay (EMD) [Babon *et al.*, 1995; Youil *et al.*, 1996] and Chemical Cleavage Analysis [Cotton *et al.*, 1988] methods. These procedures are all time-consuming and laborious, and the possibility for upscale and automation is limited, but they are still commonly used for both academic and diagnostic purposes. Current polymorphism detection methods have been designed to enable large-scale, high throughput genotyping that can be automated [Landegren *et al.*, 1998], and can be divided into the methods that are used to distinguish alleles and methods that are used to analyse the data [Gut, 2001]. Some of the methods used to distinguish alleles include hybridisation, such as the use of molecular beacons [Tyagi *et al.*, 1998]; enzymatic methods, which include the amplification refactory mutation system (ARMS) [Newton *et al.*, 1989]; TaqMan technology [Holland *et al.*, 1991; Livak *et al.*, 1995]; primer extension [Syvänen, 1999]; Flap endonuclease [Harrington & Lieber, 1994] and oligonucleotide ligation assays [Landegren *et al.*, 1988]. The results can then be analysed using one of several formats including electrophoresis, plate readers, mass spectroscopy [Gut, 2001], denaturing HPLC (DHPLC) [Underhill *et al.*, 1996] and arrays [Shuber *et al.*, 1997; Wang *et al.*, 1998]. However, although these methods can efficiently detect polymorphisms, they are not able to specifically locate them and follow-up analysis using automated sequencing [Sanger *et al.*, 1977] is required.

Unfortunately, it has been harder to locate genes and associated variants contributing to the risk and aetiology of common disease such as diabetes, heart disease, autoimmune disease, cancer and psychiatric disorders. These difficulties arise as the risk of developing complex diseases is modified and influenced by interactions between a variable number of predisposing risk alleles, the environment and an individual's lifestyle [Collins *et al.*, 1998]. By studying stretches of DNA containing SNPs associated with a disease trait, researchers have begun to reveal relevant genes associated with disease. Analysis of polymorphism in these genes helps the determination of risk distribution between populations, estimating what enhances risks within susceptible populations and possibly predicting individual risk. Evaluation of the functional impact and risks caused by these polymorphisms can provide insight into the way a particular disease or disorder works and ultimately lead to personalised

pharmacogenetics – the improved management of patients brought about by designing genotype-based medical treatment/therapy and prevention strategies.

1.2.3.2 *IN SILICO* METHODS

The sequencing of the human genome [Collins & Galas, 1993; Marshall, 1997; Venter *et al.*, 2001] and the advent of modern genomic technology allowing genome-wide analysis has fuelled a renewed interest in SNP detection and the subsequent development of *in silico* methods with which to detect them. Access to a myriad of DNA sequence information from random sets of individuals and the ability to rapidly screen a multitude of candidate genes for polymorphisms using *in silico* methods has opened the way to understanding complex disease. Many parties, both academic and industrial, have recognised the impact SNPs may have on branches of science ranging from population genetics, evolutionary biology [Cavalli-Sforza, 1998], pharmacogenomics [Evans & Relling, 1999], forensics, cancer, genetic disease research [Brookes, 1999], cloning, physical mapping [Collins, 1999] and association genetics [Kruglyak, 1999]. In response, many SNP discovery projects have been initiated, all of which are actively seeking to discover, catalogue and map SNPs in every known gene, determine allele frequencies in various populations using phenotype directed methods and determine the structural and functional impact on the protein encoded.

1.2.4 COMPUTER-BASED SEQUENCE ANALYSIS PROGRAMS

1.2.4.1 SEQUENCE DATABASES

The rapid accumulation of sequence data generated by the Human Genome Project [Boguski *et al.*, 1996; Marshall, 1997] and the large-scale sequencing of Expressed Sequence Tags (ESTs) [Adams *et al.*, 1991; Hillier *et al.*, 1996] has greatly facilitated the discovery of SNPs. The sequence data generated by these projects can be accessed through a series of sequence databases such as Genbank [Benson *et al.*, 1996], the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database

[Stoesser *et al.*, 2002] and the DNA Data Bank of Japan (DDBJ) [Tateno *et al.*, 2002], all three of which swap information daily [Benson *et al.*, 1996; Stoesser *et al.*, 2002; Tateno *et al.*, 2002]. Using these public databases to detect polymorphisms can be approached in one of two ways: by searching databases containing EST data, or by identifying sequence differences in overlapping regions between large-insert clones such as BACs (bacterial artificial chromosomes) and PACs (P1-derived artificial chromosomes) generated by the human genome project [Gu *et al.*, 1998]. The Expressed Sequence Tag database (dbEST) in particular has proved to be an extremely valuable tool that has accelerated the discovery and further study of large numbers of novel paralogous and orthologous genes and gene variants [Board *et al.*, 1997; Board, 1998; Fernández-Cañón & Peñalva, 1998; Wittenberger *et al.*, 2001; Bera *et al.*, 2002b]. Established in 1992 [Boguski *et al.*, 1993], dbEST is probably the most widely utilised sequence database and most rapidly expanding source of new sequence data corresponding to both known and unknown expressed genes. Approximately 50% of the data is generated through a collaboration between the Washington University Genome Center (St Louis) and Merck Pharmaceuticals [Benson *et al.*, 1996; Hillier *et al.*, 1996], data which is also found at the Washington University Genome Center web server http://genome.wustl.edy/gsc/gschmpg.html. The remaining sequences are submitted by various independent laboratories. The sequence data comes in the form of ESTs, 300-400 bp cDNA sequences produced by high throughput automated single pass sequencing of cDNA clones derived from different tissue libraries at three different developmental stages: fetal, infant and adult [Hillier *et al.*, 1996]. These are representative of many different individuals hence provide a means to discover polymorphism in a diverse set of samples. Normalised libraries have been constructed to reduce the number of highly expressed genes to allow better representation of rare messages [Soares *et al.*, 1994; Bonaldo *et al.*, 1996]. Error rates are particularly high at 2%-3% due to single pass sequencing and the reverse transcriptase used in this process [Adams *et al.*, 1991; Gu *et al.*, 1998]. In addition, contaminating vector sequences and other spurious sequences are sometimes submitted to the database, causing cDNA identification problems [Boguski *et al.*, 1993]. As most genes are expressed in more than one tissue type, multiple sequences representing one gene can be generated from the different libraries. There are currently 4,341,023 human ESTs derived from 426 tissue libraries in the dbEST. As this information is not curated, a number of databases designed to organise and therefore enrich the information provided by ESTs have been

developed, such as UniGene [Boguski & Schuler, 1995; Schuler *et al.*, 1996], the Genexpress Index [Houlgatte *et al.*, 1995], the Merck Gene Index [Williamson, 1999], the Sequence Tag Alignment and Consensus Knowledge (STACK) resource [Miller *et al.*, 1999] and The Institute for Genomic Research (TIGR) Human Gene Index [http://www.tigr.org]. These indices have used various procedures to group EST sequence information, full-length mRNA and expressed sequence tagged sites (eSTS) into gene clusters – sets of sequences that are likely to represent one gene. ESTs that cannot be clustered are considered new, uncharacterised genes [Boguski & Schuler, 1995].

1.2.4.2 Sequence Database Analysis Programs

Sequence databases have provided the basis for the development of computer-based programs designed for SNP discovery, and in the last five years a large number of these programs have been released. These programs encompass those that use simple alignment/assembly tools [Altschul *et al.*, 1997] and require subsequent manual filtering processes and visual scanning [Board *et al.*, 1998; Gu *et al.*, 1998; Garg *et al.*, 1999; Picoult-Newberg *et al.*, 1999], and those that use a suite of programs requiring no manual scanning [Buetow *et al.*, 1999; Marth *et al.*, 1999; Retief *et al.*, 1999].

The simpler methods are the clustering programs, which can be applied to various sequence databases and used to assemble and align the data within. Some examples are the TIGR Assembler [Sutton *et al.*, 1995], STACK_PACK [Miller *et al.*, 1999], Phrap [http://www.genome.Washington.edu/] and Basic Local Alignment Search Tool (BLAST) [Altschul *et al.*, 1997]. Of these, the BLAST alignment tool, which accesses information in dbEST and performs sequence similarity searches [Altschul *et al.*, 1997], has been widely and successfully used in the field of SNP detection [Board, 1998; Forsberg *et al.*, 1999; Blackburn *et al.*, 2000]. The disadvantage of these simple programs is that the information in the sequence databases is highly error prone due to base-calling/cDNA synthesis/cloning errors as mentioned above hence there is a possibility that any sequence variations identified will be sequencing errors. Successful use of EST database mining is dependent upon the strategies used to filter sequence

errors from the alignments [Picoult-Newberg *et al.*, 1999]. Filtering processes generally involve excluding variants that only occur once, as random errors are not likely to occur in the same position, excluding variations that occur in the number of bases in a repetitive string and excluding those surrounded by poor quality sequence.

SNP detection methods can be individually designed to meet the requirements of the researcher. These methods are developed around core programs such as Phred – a base calling program that also assesses base quality [Ewing and Green 1998; Ewing *et al.*, 1998], Phrap – a sequence alignment and contig assembly program that also provides quality assessment [http://www.genome.Washington.edu/], Consed – used to view and edit alignments and contigs and to view original traces [Gordon *et al.*, 1998], and SWAT and Cross_Match – pre-clustering steps used to search and compare sequences against a database [http://www.phrap.org/]. These programs can be used either individually or in any required combination however, depending on the programs used, the resulting alignments may still require intensive manual analysis. Different combinations of these programs have been used with success, although detection of false positives has been reported to be as high as 37%-48% [Picoult-Newberg *et al.*, 1999; Garg *et al.*, 1999]. Another program known as PolyPhred, which scans alignments for heterozygosity, has also been designed for use in conjunction with Phred, Phrap and Consed [Nickerson *et al.*, 1997]. This has been used successfully, but also requires visual scanning [Nickerson *et al.*, 1997; Nickerson *et al.*, 1998].

At the more sophisticated end of the scale, detection of SNPs in sequence databases has been fully automated by developing a suite of programs thus removing the need for manual filtering and scanning. Some examples include the FAST_PAN strategy, used to identify novel gene orthologues and paralogues [Retief *et al.*, 1999], PolyBayes [Marth *et al.*, 1999] used by the SNP consortium, and the widely used SNP Finder program [Buetow *et al.*, 1999]. Previously known as the SNPpipeline, the SNP Finder program utilises a suite of programs to detect polymorphisms in UniGene clusters. These include the Phrap and Phred programs described above, and Demiglace, which uses five filters to eliminate variations that are not likely to represent true polymorphism [Buetow *et al.*, 1999]. Each potential polymorphism is highlighted and ranked between 0.01 and 1.0, reflecting the probability that a nucleotide at this position is heterozygous. The score is influenced by the number of sequences in the alignment,

the source and diversity of the EST libraries in the UniGene cluster. Detection of false positives using this method has been reported to be relatively low at 18% [Buetow *et al.*, 1999]. Access to the original trace data of UniGene sequences is generally provided, allowing the user to further verify potentially polymorphic sequences. The user is also able to apply this program to independently generated trace data. The SNP Finder program is being successfully used by the Cancer Genome Anatomy Project-Genetic Annotation Initiative (CGAP-GAI) to compile a database of SNPs found in cancer related genes and by a number of other laboratories submitting data to dbSNP.

1.2.5 SNP DATABASES

Access to the SNPs generated by database mining strategies is required to allow further analysis and characterisation by independent laboratories interested in a specific gene or disease [Brookes, 1999]. To meet this demand, a wide range of databases has been created by various SNP detection projects. These projects are divided into two factions: the non-profit government-supported and independent groups who believe that the data generated should be freely accessible to the public, and projects supported by industry and pharmaceutical companies, such as Celera Genomics, Incyte Pharmaceuticals, Orchid Biosciences, Curagen, AlphaGene, Genset, Oxford Molecular Bioinformatics, Pangea Systems Inc. and Paracel, who believe this information should be patented, in order to secure intellectual property, and provided at a cost [Marshall, 1997; Brookes, 1999]. Both factions are competing to be the first to compile the most complete database. Despite this, the non-profit organisations have managed to make many databases available to the public in a relatively short time. These encompass those that are devoted entirely to specific loci or disease(s), which can be found at the Human Genome Organisation (HUGO) [Cotton *et al.*, 1998; now the Human Genome Variation Society (HGVS)], those listing all the known SNPs responsible for genetic disease such as the Human Gene Mutation Database [Krawczak *et al.*, 2000], those that are interested in a particular group of genes and those aiming to provide a comprehensive summary of SNPs (for extensive list of databases see Baxevanis, 2002 and the URL http://hgmp.mrc.ac.uk/GenomeWeb/human-gen-db-mutation.html). Integration of SNP data with information about the gene in which it is found, any known links to disease,

mapping information, structural information and links to relevant literature can be found at Ensembl – a joint project between EMBL and the European Bioinformatics Institute (EBI), the National Center for Biotechnology Information (NCBI), the University of California at Santa Cruz (UCSC), The SNP Consortium (TSC), Online Mendelian Inheritance in Man (OMIM) [Hamosh *et al.*, 2000], GenAtlas [Frézal, 1998], GDB (GDB™ Human Genome Database at http://www.gdb.org/) and GeneCards™ [Rebhan *et al.*, 1997].

The government funded projects have been aided by the Polymorphism Discovery Resource (PDR), which was developed through a collaborative effort between the National Human Genome Research Institute (NHGRI), the Center for Disease Control (CDC) and the National Institute of Environmental Health Sciences (NIEHS) to promote the discovery of polymorphisms in complex disease [Collins *et al.*, 1998]. The PDR provides a readily accessible, high utility cohort of cell lines and DNA samples. These samples have been collected from 450 diverse and unrelated U.S. individuals descended from one or more of the major geographic regions of the world in order to maximise the chances of discovering common DNA sequence variants. These include individuals descended from non-Hispanic white Europeans, non-Hispanic black Africans, the Americas and several Eastern and Southern Asian countries, with a 1:1 ratio of males to females [Collins *et al.*, 1998]. This cohort has been designed on the basis that common polymorphisms, which occur at high frequency, are found globally. It has been estimated that approximately 85% of the worldwide gene variation is common to all human populations, with differing allele frequencies, with the remainder being population specific [Barbujani *et al.*, 1997]. Due to potential ethical problems mainly relating to racial discrimination, this resource was not intended for population specific medical or anthropological studies, so information regarding an individuals medical and ethnic status has not been included [Marshall, 1997; Collins *et al.*, 1998; Brookes, 1999].

1.2.5.1 Databases Providing a Comprehensive Catalogue of SNPs

Three databases – the Single Nucleotide Polymorphism database (dbSNP; http://www.ncbi.nlm.nih.gov/SNP/) [Sherry *et al.*, 2001], the Human Genome Bi-Allelic Sequences database (HGBASE; http://www.hgbase.interactiva.de/) [Brookes *et al.*, 2000] and to a lesser extent the TSC (http://snp.cshl.org/) [Marshall, 1999], have aimed to provide a comprehensive overview of SNPs that is freely accessible to the public. Both the dbSNP and HGBASE were launched in 1998, dbSNP by the NCBI (U.S.A.) and NHGRI [Sherry *et al.*, 1999], and HGBASE by Interactiva GmbH and the research team of Anthony Brookes at the Karolinska Institute [Brookes *et al.*, 2000]. In November 2001, HGBASE became known as the Human Genome Variation database (HGVbase; http://hgbase.cgb.ki.se/) [Fredman *et al.*, 2002]. Although there are some differences in the way dbSNP and HGVbase are run, both have been designed as depositories for all known human cDNA and genomic DNA polymorphisms including SNPs, microsatellite repeats and small insertion/deletion polymorphisms, regardless of whether these affect function or not, and data is constantly exchanged between the two. Labs associated with the NHGRI, who use the PDR samples, are the main contributors to dbSNP, although submissions from other sources are included. Only data meeting submission information requirements is accepted and there are no requirements for a minimum allele frequency [Sherry *et al.*, 2001]. As of 6th November 2002, 4,889,801 SNPs had been submitted to dbSNP. HGVbase contains data from literature and other public genome databases in addition to accepting unique, high quality, fully validated submissions from other polymorphism discovery projects. Both known and suspected polymorphisms are listed with the corresponding annotations and as with dbSNP, there is no requirement for minimum allele frequencies. On 27th March 2002, HGVbase had a total of 1,451,426 entries, and approximately 50% of these had been mapped to the relevant gene [Fredman *et al.*, 2002]. Both databases provide information about the SNP detection method used, the sequence information surrounding the polymorphism, specific experimental conditions, frequency information and population details (if available) in order to enable further research applications. In addition, HGVbase is currently investigating the functional effects of some SNPs and links to all known references to the polymorphisms listed are provided [Brookes *et al.*, 2000; Fredman *et*

al., 2002], as are links to Genbank and EMBL. dbSNP however, does not contain any functional or phenotypic effects associated with the variations listed, but does provide links to several other databases for this purpose, and can be found at http://ariel.ucs.unimelb.edu.au:80/~cotton/guide1.htm [Sherry *et al.*, 2001]. The information in dbSNP is also integrated with other NCBI genomic data and resources such as LocusLink, which can also be used to access SNPs in dbSNP that have been mapped to a particular gene.

TSC, a collaborative effort between the Wellcome Trust and eleven pharmaceutical and technological companies set up in 1999 [Marshall, 1999], initially aimed to generate at least 300,000 SNPs and map half of these over a two-year period [Marshall, 1999]. A set of 24 ethnically diverse individuals were screened for SNPs by the Whitehead Institute, the Sanger Center and the Washington University School of Medicine, and the data was mapped and confirmed by the Cold Spring Harbour laboratories. As with dbSNP and HGVbase, information is provided about the sequence flanking the SNP, the major and minor alleles, trace data for each sequence and chromosomal mapping. As of 24th October 2002 approximately 1.8 million SNPs, over one-third the estimated number of common SNPs, had been discovered and mapped.

1.2.5.2 SNP Databases Dedicated to Specific Groups of Genes

Two independent initiatives have been established to detect SNPs in specific groups of genes. These include the Environmental Genome Project (EGP) – funded by the NIEHS, and the CGAP-GAI – initiated by the U.S. National Cancer Institute. In addition to listing all findings on their own web pages, both groups submit SNP data to dbSNP.

The EGP (http://www.niehs.nih.gov/envgenom/) was initiated in 1998 to improve understanding of human genetic susceptibility to environmental exposures through the detection and characterisation of SNPs in genes that are likely to influence the outcome of these environmental exposures. 554 candidate genes have been selected from sequence information provided by the Human Genome Project and other DNA

sequencing projects, and collaborators at the University of Utah Genome Center (UUGC; http://www.genome.utah.edu) and the University of Washington Genome Center (UWGC; http://www.genome.washington.edu/projects/egpsnps/) are in the process of resequencing these genes. Any SNPs identified are confirmed in multiple samples obtained from the PDR. SNP data generated by the UUGC, and recently UWGC, is curated in the GeneSNPs database (http://www.genome.utah.edu), which includes detailed maps of SNPs in expressed and regulatory regions and, if available, allele frequencies and genotype information. Each gene has its own "SNPcard" containing an annotated gene model that extends 10 kilobases both 5' and 3' of the expressed region. SNP location and implications are classified and exonic SNPs are further annotated. SNP data from other groups focussing on genes relevant to the EGP are also included.

The CGAP-GAI (http://cgap.nci.nih.gov/) was instigated to identify and characterise genetic variation in genes important in cancer in order to further our understanding of common cancer and its related phenotypes, and to improve detection of diagnosis and treatment of cancer. 10,000 SNPs have been identified in candidate cancer genes and depending on the validation status, these are described as Candidate (CGAPC), Validated (CGAPV), or Confirmed (CGAPCo) SNPs (http://lpg.nci.nih.gov/). Candidate SNPs are those that have been detected by the SNP Finder program with a score of at least 0.99. 75% of these are expected to represent common variation. If these are observed experimentally in eight or fewer individuals representing a diverse range of ethnic origins and disease states, the SNPs are declared Validated SNPs. However, as only eight samples are tested, only common SNPs are detected. When these SNPs have been tested in a minimum of five CEPH (Centre d'Etudes Polymorphism Humain; http://www.cephb.fr) families for Mendelian transmission and placed in genetic maps, they are classified as Confirmed SNPs.

1.2.5.3 Other SNP Databases

A multitude of other smaller databases also exist. These include those that have compiled SNP information from existing literature, such as 'A Database of Human

Genome Variation' (http://human.stanford.edu), or from the main databases described above such as refseq, which provides an index to a database that has mapped SNPs from dbSNP and the GAI onto human mRNA sequences (http://lpg.nci.nih.gov/) and Go!Poly (Gene Oriented Polymorphism Database; http://61.139.84.5/gopoly/) [Zhang *et al.*, 2001]. Others represent collaborations between researchers and companies and contribute their data to the main databases such as The Human SNP database (http://www-genome.wi.mit.edu/snp/human), which collaborates with Affymetrix and submits any SNPs identified to dbSNP [Wang *et al.*, 1998]. A list of other databases has been compiled by Baxevanis [2002] and at the URL http://hgmp.mrc.ac.uk/GenomeWeb/human-gen-db-mutation.html. Web-based tools such as SNPper (http://bio.chip.org), set up by the CHIP (Children's Hospital Informatics Program) Bioinformatics Tools [Riva & Kohane, 2001], have also been designed to facilitate searching the main SNP databases.

1.2.6 SUMMARY

The first goal of the SNP detection projects, cataloguing every SNP in every known gene, is well underway, with current estimates recording that over 1.4 million SNPs have been submitted into the public databases and mapped by the International SNP Map Working Group [The International SNP Map Working Group, 2001]. The current catalogue shows that our ability to find SNPs already surpasses our capacity to genotype them by 1000-fold, and although many approaches for genotyping have emerged over the last decade, these have not advanced at the same rate as the discovery of SNPs [Landegren *et al.* 1998; Brookes, 2002]. Databases have started to make progress with the second goal of determining allele frequencies, but none to date have determined haplotypes for linkage studies or functionally characterised these polymorphisms.

The databases of compiled SNPs contain a wealth of SNP information in various genes, and when used in combination with the various SNP detection programs available, it can be expected that many SNPs in a particular gene of interest will be isolated. Laboratory-based methods, discussed in §1.2.3.1, must then be applied to further

characterise the SNPs detected using these programs in different populations. The effects of these SNPs can then be evaluated, and in this way the link between SNPs and susceptibility to disease and drug response can be advanced.

1.3 Xenobiotic Metabolising Enzymes

1.3.1 Introduction

Drugs, natural or synthetic environmental toxins and dietary carcinogens are just a few of the exogenous xenobiotics faced by humans on a daily basis. These are capable of causing damage to cells and DNA through interactions with DNA and proteins and, if allowed to accumulate, these toxic products can result in cellular death. Just as toxic and also capable of contributing to cellular damage and disease are the endogenous reactive oxygen species (ROS). ROS are continually generated through many normal metabolic and oxidative processes such as lipid peroxidation, cellular respiration and intracellular metabolism of foreign compounds, in addition to generation through exogenous sources such as UV exposure and excessive iron salts [Hayes & McLellan, 1999]. Accumulation of ROS results in a cytotoxic state known as oxidative stress, a state that can cause serious cellular damage through targeting DNA, proteins and lipids. Accumulation of excessive amounts of ROS are responsible for the majority of cellular damage and subsequent production of hydroperoxides, alkenals and catechol products, all of which can contribute to premature aging and disease such as cancer, arthritis and neurodegenerative disorders [Beckman & Ames, 1998].

In order to prevent damage caused by toxic insults, various strategies involving sequestering molecules, scavenging molecules and antioxidant enzymes and molecules have evolved [Beckman & Ames, 1998; Sheehan *et al.*, 2001]. Three groups of enzymes in particular have evolved as a defense mechanism against drug and xenobiotic detoxification: the Phase I, Phase II and Phase III enzymes. The Phase I enzymes, consisting of superoxide dismutase, catalase, thioredoxin peroxidase and predominantly the microsomal cytochrome P450 monooygenases, are responsible for the initial oxidation of the exogenous or endogenous toxic compounds. In some instances this step activates the toxic compound, hence the resulting metabolites can actually be more reactive than the parent compound and require further detoxification. This step is carried out by the Phase II enzymes, such as the glutathione *S*-transferases (GSTs), glutathione peroxidases, uridine diphosphate (UDP) -glucuronyl transferases, sulfo-transferases, epoxide hydrolase, quinone reductase and aldo-keto reductases, which

catalyse conjugation, reduction or peroxidation reactions [Hayes & Pulford, 1995; Hayes & McLellan, 1999]. Generally, Phase II enzymes catalyse the conjugation of activated xenobiotics to an endogenous water-soluble substrate such as glutathione (GSH), UDP-glucuronic acid or glycine [Sheehan *et al.*, 2001] in order to solubilise the compound and thereby facilitate further metabolism. Subsequent excretion from the cell into urine or bile [Hayes & Pulford, 1995; Seidegård & Ekstrom, 1997] is achieved through the Phase III transporters such as the multidrug-resistance-associated protein family (MRP), which consists of nine genes [Cole *et al.*, 1992; Taniguchi *et al.*, 1996; Kool *et al.*, 1997; Kool *et al.*, 1999; Bera *et al.*, 2001; Hopper *et al.*, 2001; Bera *et al.*, 2002a].

1.3.1.1 The Glutathione Transferases

The Phase II GST enzymes are a major component of cellular defence against xenobiotics in many species [Strange *et al.*, 2000; Sheehan *et al.*, 2001]. The GSTs can be divided into two superfamilies: the cytosolic or soluble GSTs and the microsomal GSTs. The microsomal GSTs, known as the Membrane Associated Protein in Eicosanoid and Glutathione metabolism (MAPEG) family, consist of six genes. These are primarily involved in detoxification of leukotriene A_4, an epoxide product of arachidonic acid metabolism [Morgenstern *et al.*, 1982; Jakobsson *et al.*, 1999]. The cytosolic GST superfamily are predominantly involved in the detoxification of a diverse range of toxic xenobiotics and endogenous toxins [Hayes & Pulford, 1995]. Using the reduced tripeptide GSH, the GSTs are able to catalyse nucleophilic aromatic substitutions, reversible Michael additions to α,β-unsaturated ketones and epoxide ring-opening reactions, all of which result in the formation of thioether conjugates [Armstrong, 1997]; catalyse the reduction of hydroperoxides which results in the formation of oxidised glutathione (GSSG) [Hayes & McLellan, 1999]; and catalyse isomerisations [Burgess *et al.*, 1989]. In general, the reactions render the electrophiles less reactive by converting them to more hydrophilic compounds. In some instances however, the toxicity of some substrates is increased. Conjugation may render the compounds more reactive as occurs with halogenated alkanes or alkenes, which are nephrotoxic in either the conjugated or activated form, and therefore require further

detoxification [Anders *et al.*, 1988]. In some instances, the detoxification reaction is reversible and the original compound is regenerated [Atsmon *et al.*, 1990b; Ploemen *et al.*, 1994]. These toxifying properties can be exploited in cancer chemotherapy to treat tumours that over-express GSTs [Hayes & Pulford, 1995].

The remainder of this introduction will focus exclusively on the cytosolic GST family of enzymes and the MAPEG family will not be discussed further.

1.3.2 Properties of the GSTs

1.3.2.1 Members of the GST Superfamily

Eight human cytosolic/soluble classes of GST have been identified: Alpha (GSTA), Kappa (GSTK), Mu (GSTM), Omega (GSTO), Pi (GSTP), Sigma (GSTS), Theta (GSTT) and Zeta (GSTZ) [Mannervik *et al.*, 1985; Meyer *et al.*, 1991; Hussey & Hayes, 1992; Pemble *et al.*, 1996; Board *et al.*, 1997; Kanaoka *et al.*, 1997; Board *et al.*, 2000], each of which may include up to five separate isoforms. Although these are collectively known as the cytosolic GSTs, *GSTA4* and the Kappa class GST are located in the mitochondria [Pemble *et al.*, 1996; Gardner & Gallagher, 2001]. Classification of the GST classes is based on the relationship between primary structures with emphasis on the highly conserved N-terminal region, tertiary and quaternary structural properties, immunological properties and kinetic properties such as substrate and inhibitor specificities [Mannervik *et al.*, 1985; Sheehan *et al.*, 2001]. Members within a class generally have greater than 40% amino acid similarity and tend to cluster together at single loci [Hayes & Pulford, 1995; Webb *et al.*, 1996], whereas between classes members possess less than 30% amino acid similarity [Hayes & Pulford, 1995]. It has been proposed that the Kappa GST gene may have been the evolutionary ancestor of the Theta GST gene [Pemble *et al.*, 1996; Armstrong, 1997], from which the Alpha, Mu, Omega, Pi, Sigma and Zeta GST genes derived through gene duplication events hence allowing organisms to adapt to various toxic stresses during the course of evolution [Pemble & Taylor, 1992; Armstrong, 1997; Sheehan *et al.*, 2001].

The Alpha, Mu and Pi GSTs were originally identified in rat liver cytosol through their activity with aryl halides, especially the model GST substrate 1-chloro-2,4-dinitrobenzene (CDNB) [Clark *et al.*, 1973] and their ability to bind to GSH affinity matrices [Mannervik *et al.*, 1985]. However, these methods failed to detect all of the presently known GSTs, either because they are not highly active with CDNB or because they do not bind to GSH affinity matrices. For example, the Theta class GSTs were discovered as a result of their capacity to utilise novel substrates such as 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP), *p*-nitrophenethyl bromide (*p*NPB) and *p*-nitrobenzyl chloride (*p*NBC) [Meyer *et al.*, 1991], and the Kappa class GST was originally isolated from the rat mitochondrial matrix [Harris *et al.*, 1991]. Database analysis, which identifies genes on the basis of sequence similarity rather than substrate specificity, has been responsible for the discovery of new GST families including the Omega and Zeta class GSTs [Pemble *et al.*, 1996; Board *et al.*, 1997; Board *et al.*, 2000], and new isoenzymes in existing families such as the *GSTA4* gene [Board, 1998]. With the continual expansion of sequence databases, there is a high likelihood that other novel GST families and isoenzymes will be discovered.

Of all the GST classes, the Alpha and Mu class GSTs appear to be the most diverse, with at least five distinct genes found in mammals. Five alpha class genes have been described in humans: *GSTA1* [Rozen *et al.*, 1992], *GSTA2* [Klöne *et al.*, 1992; Röhrdanz *et al.*, 1992], *GSTA3* [Suzuki *et al.*, 1993], *GSTA4* [Desmots *et al.*, 1998] and *GSTA5* [Morel *et al.*, 2002], and are clustered within a 300 Kb region on chromosome 6p.12.1p12-2 [Board & Webb, 1987; Morel *et al.*, 2002]. All but *GSTA5* have been cloned [Tu & Qian, 1986; Board & Pierce, 1987; Board & Webb, 1987; Rhoads *et al.*, 1987; Board, 1998; Hubatsch *et al.*, 1998; Liu *et al.*, 1998; Johansson & Mannervik, 2001a]. Although the gene sequence for *GSTA5* appears to contain a complete coding region sequence and no obvious defects have been observed to suggest it is non-functional, cDNA sequences have not yet been detected in the EST database implying either very low expression levels, or specific tissue distribution [Morel *et al.*, 2002]. *GSTA1*, *GSTA2* and *GSTA3* are all similar with >90% sequence identity. *GSTA4* however, shares only 25% sequence identity with these three genes [Board, 1998]. Seven pseudogenes are also located in this cluster, which is ordered as follows: 5'-*GSTA4-GSTAP4-GSTA3-GSTAP2-GSTAP3-GSTA5-GSTAP1-GSTA1-GSTAP5-GSTA2-GSTAP6-GSTAP7*-3' [Morel *et al.*, 2002]. The pseudogenes can be distinguished from

the functional genes by any combination of the following: deletion of single nucleotides or complete exons, cryptic stop codons and irregular splicing signals [Klöne *et al.*, 1992; Suzuki *et al.*, 1993; Lörper *et al.*, 1996].

The Mu class is also made up of five closely linked isoforms, all of which have been cloned and characterised as recombinant proteins. These are *GSTM1* [DeJong *et al.*, 1988; Seidegård *et al.*, 1988], *GSTM2* [Vorachek *et al.*, 1991], *GSTM3* [Campbell *et al.*, 1990], *GSTM4* [Comstock *et al.*, 1993; Ross & Board, 1993; Zhong *et al.*, 1993a] and *GSTM5* [Takahashi *et al.*, 1993]. All five isoforms have been mapped to a 20 Kb cluster on chromosome 1p13.3 [Pearson *et al.*, 1993; Ross *et al.*, 1993]. With the exception of *GSTM3*, which is arranged tail-to-tail with *GSTM5*, the Mu class genes are oriented head-to-tail to each other as follows: 5' - *GSTM4-GSTM2-GSTM1-GSTM5-GSTM3* - 3' [Xu *et al.*, 1998b; Patskovsky *et al.*, 1999a]. There have been reports of GST Mu-like sequences mapping to chromosomes 3 [Islam *et al.*, 1989], 6 [DeJong *et al.*, 1991; Ross *et al.*, 1993] and 13 [DeJong *et al.*, 1991], however these have not been further characterised and there is speculation as to whether these are related Mu class genes or pseudogenes [DeJong *et al.*, 1991; Ross *et al.*, 1993]. The five isoforms are highly similar, especially within the N-terminal region [Rowe *et al.*, 1997], with *GSTM1*, *GSTM2*, *GSTM4* and *GSTM5* sharing more than 84% sequence identity with each other. *GSTM3* is the exception, sharing less than 75% sequence identity with the other Mu isoforms [Ross & Board, 1993; Patskovsky *et al.*, 1999a], due to extensions found at either end of the protein. It possesses four extra amino acids at the acetylated N-terminus and an extra three at the C-terminus [Campbell *et al.*, 1990; Rowe *et al.*, 1997]. In light of these differences, it has been suggested that the Mu class GSTs can be divided into two groups: the M1-like and the M3-like sequences [Patskovsky *et al.*, 1999a]. *GSTM1*, *GSTM2*, *GSTM4* and *GSTM5* are believed to have originated from the M1-like gene through recombination and gene conversion events, whilst the *GSTM3*-like gene, now the *GSTM3* gene, remained unchanged [DeJong *et al.*, 1991; Taylor *et al.*, 1991; Vorachek *et al.*, 1991].

The remaining GST classes are less diverse. Database analysis has recently revealed the existence of two Omega class isoforms: *GSTO1* and *GSTO2* [Board *et al.*, 2000; Morel *et al.*, 2001; Whitbread *et al.*, In Press], both of which have been cloned and are currently being characterised [Morel *et al.*, 2001; Whitbread *et al.*, In Press]. These two

isoforms share 64% sequence identity, and have been mapped to chromosome 10q24.3 [Morel *et al.*, 2001; Whitbread *et al.*, In Press]. A third member of the GST Omega class, *GSTO3*, has been located on chromosome 3 [Whitbread *et al.*, In Press]. Although this shows 91% sequence homology to *GSTO1*, the lack of introns and different chromosomal location imply that this is a pseudogene. *GSTO1* has also recently been identified as a monomethylarsonic acid (MMA^V) reductase [Zakharyan *et al.*, 2001].

Two human Theta class GSTs sharing 55% amino acid identity have been cloned and characterised: *GSTT1* [Meyer *et al.*, 1991; Pemble *et al.*, 1994] and *GSTT2* [Hussey & Hayes, 1992; Tan *et al.*, 1995; Coggan *et al.*, 1998]. *GSTT1* and *GSTT2* are arranged in a tail-to-tail orientation on chromosome 22q11-2 and are separated by a 50 Kb stretch of DNA [Tan *et al.*, 1995; Webb *et al.*, 1996; Coggan *et al.*, 1998]. *GSTT2* is found head-to-head with the D-dopachrome tautomerase gene (*DDCT*). The two are believed to be divided by a bidirectional promoter [Coggan *et al.*, 1998]. In addition, a duplication of the region encoding the *DDCT* gene and a *GSTT2* pseudogene (*GSTT2P*) is found upstream of the *GSTT1* and *GSTT2* genes [Coggan *et al.*, 1998].

Unlike the GST families described above, only one isoenzyme has been isolated for the Pi, Zeta, Kappa and Sigma class GSTs. *GSTP1* has been cloned, extensively characterised [Kano *et al.*, 1987; Cowell *et al.*, 1988; Board *et al.*, 1989; Morrow *et al.*, 1989] and mapped to chromosome 11q13 [Moscow *et al.*, 1988; Board *et al.*, 1989]. A *GSTP1* reverse transcribed pseudogene has also been reported and mapped to chromosome 12q13-14 [Board *et al.*, 1989; Board *et al.*, 1992]. *GSTZ1* was identified using EST sequence alignments and phylogenetic studies [Board *et al.*, 1997] and has been mapped to chromosome 14q24.3 [Blackburn *et al.*, 1998; Fernández-Cañón *et al.*, 1999]. This gene has also been identified as a maleylacetoacetate isomerase [Fernández-Cañón & Peñalva, 1998]. A human *GSTK1* cDNA was identified through EST database analysis [Pemble *et al.*, 1996], using the previously identified rat Kappa class GST [Harris *et al.*, 1991; Pemble *et al.*, 1996], however no reports on the further characterisation of human *GSTK1* are available. The human *GSTS1*, a class originally defined through sequence alignments of S-crystallin and GST cDNAs from molluscs [Buetler & Eaton, 1992], was identified as a homologue of the rat prostaglandin (PG) D

synthase (PGDS) and has been mapped to chromosome 4q21-22 [Kanaoka *et al.*, 1997; Kanaoka *et al.*, 2000].

1.3.2.2 GST Structure and Substrate Specificities

The GSTs are able to detoxify a vast number of xenobiotics and each class has distinct preferences for different substrates. A common substrate able to be utilised by every class of GST has not been identified. Although there is some overlap, even members within a class no matter how similar, can have distinct substrate specificities and catalytic efficiencies. This is possibly due to variations in the active site residues that exert slight structural differences [Chow *et al.*, 1988]. For example, of the eleven amino acids that differ between the GSTA2-2 and GSTA1-1 proteins, four are located in the active site. It has been proposed that an accumulation of single amino acid residue replacements may be an important mechanism for generating diversity within a class for various substrates [Chow *et al.*, 1988]. Despite these differences, all soluble GSTs appear to have a generally similar structure [Sheehan *et al.*, 2001].

1.3.2.2.1 GST Structure

The first three-dimensional structure of a human GST was solved in 1992 for the Pi class GST [Reinemer *et al.*, 1992]. Subsequently, representative structures for the human Alpha, Mu, Omega, Pi, Theta and Zeta GSTs, and a rodent Sigma class GST have been determined [Sinning *et al.*, 1993; Raghunathan *et al.*, 1994; Kanaoka *et al.*, 1997; Rossjohn *et al.*, 1998; Bruns *et al.*, 1999; Board *et al.*, 2000; Polekhina *et al.*, 2001]. The structure for the Kappa class GST has not been determined. All GSTs have the same fundamental structure, as the changes that distinguish each class are relatively conservative and do not dramatically alter the structure [Dirr *et al.*, 1994]. Each GST exists as a dimer with one catalytic site per subunit, each of which consists of two domains connected by short linker regions: the N-terminal domain I and the C-terminal domain II [Dirr *et al.*, 1994]. The active site appears as a cleft along the domain interface. Although catalytically independent, a fully functional active site requires

both subunits [Danielson & Mannervik, 1985]. The N-terminal domain I, loosely referred to as the GSH binding site or G-site [Mannervik *et al.*, 1978], is composed of the first 80 or so residues arranged in four β-sheets flanked by three α-helices and provides the binding site for GSH [Armstrong, 1997; Sheehan *et al.*, 2001]. It also contributes some of the contacts for substrate binding. This domain contains a catalytically essential tyrosine, serine or cysteine residue that interacts with the thiol group of GSH [Sheehan *et al.*, 2001]. The C-terminal domain II, known as the hydrophobic substrate-binding site or H-site [Mannervik *et al.*, 1978], is generally helical in structure, although the number of helices varies between GST classes [Sheehan *et al.*, 2001]. Domain II is composed of the remaining amino acid residues and makes most of the contacts for the H-site mainly along the face of the α4 helix and the C-terminal tail and also contributes a highly conserved aspartic acid residue from the α4 helix to the G-site [Armstrong, 1997; Sheehan *et al.*, 2001]. A small number of residue differences in this domain are responsible for the dramatic differences in substrate specificity observed between the GSTs, as these are able to regulate the hydrophobicity and steric limitations of the H-site [Danielson *et al.*, 1987; Wilce & Parker, 1994]. Despite these changes, specificity for GSH is retained [Wilce & Parker, 1994].

Many features of the subunit interface, which diverges to create a V-shaped crevice, are conserved between the GST classes [Sinning *et al.*, 1993; Dirr *et al.*, 1994; Armstrong, 1997], but the differences that do exist among subunits of different classes, and even within classes, make their interfaces incompatible, as these alter the electrostatic interactions contributing to the stability and function of the protein. These differences have been proposed to influence formation of heterodimers, which require highly similar subunit interfaces. Only the GSTA1/A2 and GSTM1a/M2 heterodimers have been observed *in vivo* [Laisney *et al.*, 1984; Stockman *et al.*, 1985; Tsuchida *et al.*, 1990; Hussey *et al.*, 1991]. GSTA1/A4 and GSTM2/P1 heterodimers have been observed *in vitro*, however their existence *in vivo* is not possible due to subunit interface incompatibility [Pettigrew & Colman, 2001; Gustafsson *et al.*, 2002]. The interface of the Alpha, Mu, Pi and Zeta class GSTs is characterised by a key and lock feature in which an aromatic residue (phenylalanine for Alpha, Mu and Pi; methionine for Zeta) acts as a key extending from the loop between the α2-helix and β3 sheet of domain I on

one subunit, and fits into the hydrophobic lock provided by helices α4 and α5 in domain II of the other subunit [Armstrong, 1997]. This particular interaction is not seen in the Theta and most Sigma class GSTs, hence these have a relatively flat and more hydrophilic interface [Ji *et al.*, 1995; Armstrong, 1997; Stevens *et al.*, 1998; Stevens *et al.*, 2000]. One mammalian Sigma class GST however, the haematopoietic prostaglandin D synthase, does have a lock and key motif and a distinct active site cleft [Kanaoka *et al.*, 1997].

Unique structural differences exist between the different GST classes in spite of similar overall topologies due to substitutions, deletions and insertions in the β-sheets and α-helices [Dirr *et al.*, 1994]. The most notable are the Alpha class GSTs, which are distinguished by the presence of an extra helix, the α9 helix (residues 210 to 220), in the C-terminal domain II, which forms a flexible cap over the H-site and is able to adopt different conformations depending on the substrate present [Sinning *et al.*, 1993; Gustafsson *et al.*, 1999]. This creates a smaller and more strongly hydrophobic site compared to the Mu and Pi classes [Sinning *et al.*, 1993]. The Mu class GSTs possess what has been denoted a "mu loop" [Ji *et al.*, 1992], which is located between β2 and α2 and results in a deeper active site cleft than that found in the Pi class [Ji *et al.*, 1992; Raghunathan *et al.*, 1994; Wilce & Parker, 1994]. GST Omega is distinguished by a 19-residue proline rich N-terminal extension and seven α-helices in the C-terminal domain II. The active site is considerably more open than observed in other GSTs and the H-site is relatively large and polar [Board *et al.*, 2000]. A consensus pattern of residues unique to the Theta class GSTs concentrated in the α2 helix has been observed [Rossjohn *et al.*, 1996] and GSTT2-2 possesses a long C-terminal extension, which blocks both the H-site and most of the G-site [Rossjohn *et al.*, 1998]. Zeta has a very small, almost buried polar, hydrophilic active site and lacks the V-shaped interface [Polekhina *et al.*, 2001].

The most fundamental difference between the GST classes is the catalytically essential residue in domain I that interacts with the sulfur atom of GSH. The Alpha, Mu, Pi and Sigma class GSTs all possess a catalytically essential tyrosine [Wilce & Parker, 1994; Kanaoka *et al.*, 1997; Polekhina *et al.*, 2001]. In contrast, the Omega class GST uses a cysteine residue [Board *et al.*, 2000], and the Theta and Zeta classes a serine residue

[Board *et al.*, 1995; Wilce *et al.*, 1995; Polekhina *et al.*, 2001]. In addition, there is evidence for the involvement of residue Arg15 in the Alpha class [Stenberg *et al.*, 1991a; Björnstedt *et al.*, 1995], His108 in GSTM1-1, S108 in GSTM4-4 [Patskovsky *et al.*, 1999b], R108 in GSTM2-2, R112 in GSTM3-3 [Patskovsky *et al.*, 2000], R100 in GSTT2-2 [Caccuri *et al.*, 2001] and possibly Arg100 in GSTP1-1 [Patskovsky *et al.*, 2000] in the formation and stabilisation of the thiolate anion.

1.3.2.2.2 GST Substrate Specificities

1.3.2.2.2.1 Detoxification Reactions

Each class of GST is able to catalyse reactions with specific substrates due to their unique hydrophobic binding sites hence is able to provide protection against the toxic effects of a large number of commonly encountered and harmful substrates with various functional groups. These include α,β-unsaturated aldehydes produced endogenously by lipid peroxidation [Esterbauer *et al.*, 1991] or environmentally through industrial waste, tobacco smoke and fuel exhaust [Beauchamp *et al.*, 1985]; the base propenals generated by oxidative damage to DNA [Alarcon & Meienhofer, 1971; Alarcon, 1976]; or the *o*-quinones, the oxidation products of catecholamines such as aminochrome, dopachrome and noradrenochrome [Baez *et al.*, 1997].

1.3.2.2.2.1.1 GST Alpha

The Alpha class GSTs, particularly GSTA1-1 and GSTA2-2, possess high selenium-independent glutathione peroxidase activity (GPx-II), which allows them to reduce cellular peroxides such as fatty acid hydroperoxides and phospholipid hydroperoxides to the corresponding alcohols [Awasthi *et al.*, 1980; Singhal *et al.*, 1992a]. Although several glutathione peroxidases (GPx) have been identified in the liver and the specific activity of GSTA1-1 and GSTA2-2 is much lower than the selenium-dependent glutathione peroxidases (GPx-I), the total amount of GST protein in the liver is much higher than the other GPx proteins, suggesting that the GSTs play a major role in the protection mechanisms against lipid peroxidation [Hayes & McLellan, 1999; Zhao *et*

al., 1999]. Cholesterol A-oxide generated by oxidation of membranes is another potentially important epoxide detoxified by the Alpha class GSTs through combination with GSH [Meyer & Ketterer, 1982; Ansari & Smith, 1990; Hayes & McLellan, 1999].

The Alpha class GSTs, GSTA4-4 in particular, are also capable of conjugating α-β-unsaturated aldehydes, the extremely reactive and genotoxic breakdown products of peroxidation reactions [Esterbauer *et al.*, 1991; Board, 1998; Hubatsch *et al.*, 1998]. These may be derived endogenously and include 4-hydroxynon-2-enal (4HNE) [Esterbauer *et al.*, 1991] and acrolein, the activated metabolic product of the cytostatic drug cyclophosphamide [Alarcon & Meienhofer, 1971]. They may also be exogenously derived and include chemicals such as crotonaldehyde and again, acrolein, that are found in industrial waste pollutants, tobacco smoke, gasoline and diesel exhaust [Beauchamp *et al.*, 1985].

1.3.2.2.2.1.2 GST Mu

The Mu class GSTs, particularly GSTM1-1, are distinguished by their ability to detoxify epoxide metabolites, which are generated by the breakdown of polycyclic aromatic hydrocarbons (PAH) and carcinogens such as Aflatoxin B_1 by Phase I enzymes. These include the PAH metabolites styrene 7,8-oxide, trans stilbene-oxide, benzo[a]pyrene 4,5-oxide, benzo[*a*]pyrene-7,8-diol-9,10-epoxide, benz[*a*]anthracene, 7-methyl-benz[*a*]anthracene, 7,12-dimethylbenz[*a*]anthracene and 3-methylcholenthrene, and the aflatoxin B_1 metabolite aflatoxin B_1-8,9-*exo*-epoxide (AFB_1) [Warholm *et al.*, 1983; Hayes & Pulford, 1995; Johnson *et al.*, 1997]. These compounds are considered highly mutagenic and oncogenic and if allowed to bind to DNA, thereby forming PAH-DNA adducts, they are able to modify the structure and function of the DNA [Pelkonen *et al.*, 1980]. These compounds are found in products of combustion such as cigarette smoke, car exhaust fumes and burnt foods, or in the case of aflatoxins, mouldy foods.

Peroxidation is not limited to fatty acid and phopholipid hydroperoxides, but also includes peroxidation of the cyclic *o*-quinones aminochrome, dopachrome, noradrenochrome and adrenochrome – oxidation products of the catecholamines dopamine, dopa, noradrenaline and adrenaline respectively [Baez *et al.*, 1997].

GSTM1-1, and GSTM2-2 in particular, possess 100 to 1000 times the activity of all the GSTs with the cyclic *o*-quinones. It has been proposed that the oxidation of catecholamines, which can produce excessive numbers of ROS, contributes to Parkinson's Disease (PD). This is believed to occur via the cyclic pro-oxidant pathway in which *o*-quinones are reduced and subsequently reoxidised. Excessive numbers of ROS are generated by this redox cycling, leading to oxidative stress, toxicity, apoptosis and ultimately neurodegeneration [Baez *et al.*, 1995]. This is usually inhibited by the products of a competing reaction in which *o*-quinones are conjugated to GSH by GSTs [Segura-Aguilar *et al.*, 1997]. More recently, it has been found that GSTM2-2 can actually catalyse the conjugation of dopamine and dopa to GSH, thereby preventing the formation of their respective cyclic *o*-quinones and subsequent redox cycling [Dagnino-Subiabre *et al.*, 2000].

1.3.2.2.2.1.3 GST Pi

GSTP1-1 has relatively high activity with the diuretic ethacrynic acid (EA), but its characteristic feature is high activity towards the α,β-unsaturated aldehydes acrolein and crotonaldehyde and against the base propenals [Berhane *et al.*, 1994]. The latter are products of DNA degradation formed during oxygen-linked radical reactions in the presence of the chemotherapeutic drug bleomycin [Giloni *et al.*, 1981; Grollman *et al.*, 1985] or γ-irradiation [Janicek *et al.*, 1985]. GSTP1-1 possesses high catalytic efficiency with the carcinogenic diol epoxides of PAHs such as such as anti-1,2-dihydroxy-3,4-oxy-1,2,3,4-tetrahydrochrysene (anti-CDE) formed from the environmental pollutant chrysene [Hu *et al.*, 1997a] and (±)-7β,8α-dihydroxy-9α,10α-oxy-7,8,9,10-tetrahydrobenzo(a)pyrene (±-anti-BPDE), the carcinogenic activated metabolite of benzo(a)pyrene, which is abundant in cigarette smoke [Robertson *et al.*, 1986; Hu *et al.*, 1997b; Sundberg *et al.*, 1998a; Sundberg *et al.*, 1998b].

1.3.2.2.2.1.4 GST Omega

Due to the presence of the active site cysteine, GST Omega does not catalyse the GST reactions typical of many GSTs, which require an active site tyrosine or serine [Stenberg *et al.*, 1991b; Board *et al.*, 1995; Board *et al.*, 2000]. As yet, a unique substrate has not been identified [Board *et al.*, 2000], although weak thiol transferase and dehydroascorbate reductase activity characteristic of the glutaredoxins has been observed [Board *et al.*, 2000]. GSTO1-1 has recently been ascribed a role in the pathway involved in the biotransformation of inorganic arsenic, a carcinogenic contaminant in drinking water in some regions of the world [Smedley & Kinniburgh, 2001; Zakharyan *et al.*, 2001]. In the penultimate step of this biotransformation pathway, also the rate-limiting step, GSTO1-1 catalyses the GSH-dependent reduction of monomethylarsonic acid (MMA^{V}) to monomethylarsonous acid (MMA^{III}) [Zakharyan & Aposhian, 1999; Zakharyan *et al.*, 2001].

1.3.2.2.2.1.5 GST Theta

The Theta class GSTs catalyses the conjugation of several important small dihaloalkenes. GSTT1-1 in particular is unusual in that it is able to either toxicate or detoxicate substrates, depending on the nature of the substrate. It detoxifies monohaloalkanes and ethylene oxide, but toxifies dihalomethanes such as dichloromethane (DCM), dihaloethanes such as dibromoethane (DBE) and other bifunctional alkylating agents in GSH-dependent reactions [Ketterer & Christodoulides, 1994; Thier *et al.*, 1996; Sherratt *et al.*, 1997]. DCM is found in paint strippers, aerosol preparations and is involved in the synthesis of plastics and pharmaceutical drugs. GSH conjugates of the dihalomethanes are believed to be mutagenic [Wheeler *et al.*, 2001b], as studies have shown that mice, which express five-fold more GSTT1-1 in liver and lungs and have a much higher catalytic ability with DCM than humans, develop liver and lung tumours when exposed to DCM. Although the health risk in humans is unknown, it is believed to be relatively low when compared to the mouse due to the lower expression and lower capacity of GSTT1-1 to activate DCM [Wong *et al.*, 1982; Burek *et al.*, 1984; Sherratt *et al.*, 2002]. DBE and 1,2,3,4-diepoxybutane are used in fumigants, anti-knock mixtures and in the rubber industry. These compounds are

converted by GSTT1-1 to mutagenic half mustards and 2-haloacetaldehydes that are capable of interacting with DNA to form miscoding adducts [Barrio *et al.*, 1972; Ozawa & Guengerich, 1983; Guengerich & Persmark, 1994; Thier *et al.*, 1996; Hayes & Strange, 2000; Wheeler *et al.*, 2001a].

In contrast, GSTT2-2 lacks activity with the alkyl halides. Its role is largely unknown, however it is reactive with secondary lipid peroxidation products such as the *trans*,alk-2-enals and *trans,trans*,alka-2,4-dienals, implying a role against lipid peroxidation and ROS [Tan & Board, 1996]. GSTT2-2 is also specifically active with the sulfate ester menapthyl sulfate, suggesting that it may be active against other sulfate esters derived from carcinogenic arylmethanols [Tan & Board, 1996]. The rat orthologue of human GSTT2-2 has been shown to metabolise sulfate esters derived from 5-hydroxymethyl chrysene and 7,12-dihydroxymethylbenz[a]anthracene [Hiratsuka *et al.*, 1990]. This activity could potentially provide protection against hepatocarcinoma development [Okuda *et al.*, 1986; Watabe *et al.*, 1986; Hiratsuka *et al.*, 1990; Tan & Board, 1996].

1.3.2.2.2.1.6 GST Zeta

GST Zeta is unique in that it is able to catalyse the metabolism of α-halopropionic acids and dihaloacetic acids [Tong *et al.*, 1998a]. In particular, GSTZ1-1 is responsible for catalysing the metabolism of the dihaloacetic acid dichloroacetate (DCA) to glyoxalate [Tong *et al.*, 1998b]. DCA is a common contaminant in drinking water [Uden & Miller, 1983] and has been shown to be hepatocarcinogenic in rodents [Bull *et al.*, 1990; DeAngelo *et al.*, 1996] in addition to exerting reproductive and embryotoxic effects [Linder *et al.*, 1995; Hunter *et al.*, 1996; Linder *et al.*, 1997]. However, DCA is not believed to be carcinogenic in humans, and is used clinically in the management of lactic acidosis and its use as neuroprotective agent has been proposed [Stacpoole, 1989]. GSTZ1-1 is irreversibly inactivated by DCA and other fluorine lacking dihaloacetic acids in rats, mice and humans by a mechanism in which the xenobiotic is biotransformed into a reactive metabolite that alkylates Cys16, located in the active site, and Cys205, located near the active site. However, GSTZ1-1 is not inactivated by α-monohaloacids and fluorine containing dihaloacetic acids [Anderson *et al.*, 1999; Wempe *et al.*, 1999; Tzeng *et al.*, 2000; Lantum *et al.*, 2002]. Inactivation of GSTZ1-1

by DCA has been shown to cause an accumulation of maleylacetone (MA), a byproduct of the tyrosine metabolism pathway [Cornett *et al.*, 1999; Tzeng *et al.*, 2000].

1.3.2.2.2.1.7 GST Kappa and GST Sigma

To date, unique electrophilic substrates have not yet been identified for either GSTK1-1 or GSTS1-1. However, the rat Kappa class GST has been found to possess relatively weak activity with CDNB and EA [Harris *et al.*, 1991], and human GSTS1-1 has been found to possess some conjugating and peroxidase activities with some typical GST model substrates [Jowsey *et al.*, 2001].

1.3.2.2.2.2 Non-detoxification Reactions

In addition to their role as detoxification enzymes, some GST isoforms play important roles in different biological pathways. The predominant non-detoxification role of the GSTs is the ability to catalyse isomerisation reactions, the most common of which are those that convert *cis* double bonds to the *trans* configuration. GSTP1-1 acts as a retinoid isomerase, converting 13-*cis*-retinoic acid (13-cRA) to all-*trans*-retinoic acid (t-RA) in a GSH-independent reaction, hence may have a significant role in regulating the biological activities of 13-cRA. Isomerisation of 13-cRA has been proposed to be an activation mechanism for the potent therapeutic and teratogenic effects of 13-cRA [Chen & Juchau, 1998]. GSTZ1-1, identified as a maleylacetoacetate isomerase, is responsible for catalysing the GSH-dependent *cis-trans* isomerisation of maleylacetoacetate (MAA) to the alkylating agent fumarylacetoacetate (FAA). This is the penultimate step in the tyrosine and phenylalanine catabolic pathway [Blackburn *et al.*, 1998; Fernández-Cañón & Peñalva, 1998]. Deficiencies in this pathway have led to diseases including alkaptonuria, phenylketonuria and several forms of tyrosinaemia, such as hereditary tyrosinaemia type I [Lindblad *et al.*, 1977; Fernández-Cañón & Peñalva, 1998]. Deficiency in MAA isomerase, or the Zeta class GST, is believed to be toxic and although this has not yet been reported, cases in which GSTZ1-1 activity is very low have resulted in severe tyrosinaemic disorder [Berger *et al.*, 1988].

GSTs are also capable of catalysing positional isomerisation of double bonds. The best characterised is that in which endogenous Δ^5-3-ketosteroids such as Δ^5-androstene-3, 17-dione and Δ^5-pregnane-3, 20-dione are converted by GSTA3-3 to Δ^4-androstene-3, 17-dione and Δ^4-pregnane-3, 20-dione, the immediate precursors of testosterone and progesterone, respectively [Johansson & Mannervik, 2001a]. The biosynthesis of prostaglandins represents yet another type of isomerisation reaction, in which an endoperoxide is converted to a β-hydroxy ketone. GSTA1-1 and GSTA2-2 catalyse the isomerisation of Prostaglandin H_2 (PGH_2) to PGD_2, PGE_2 and $PGF_{2\alpha}$ [Burgess *et al.*, 1989]. PGD_2 is also produced through isomerisation of PGH_2 by GSTS1-1, a haematopoietic or GSH-dependent PGDS [Kanaoka *et al.*, 1997], and PGE_2 is also generated by the isomerisation of PGH_2 catalysed by the Prostaglandin E synthases GSTM2-2 and GSTM3-3 [Beuckmann *et al.*, 2000]. PGD_2 and PGE_2 are unstable molecules that undergo hydration to form the PGJ_2 series and PGA_2 respectively, which inhibit cell proliferation [Kato *et al.*, 1986; Atsmon *et al.*, 1990a; Kim *et al.*, 1993]. As these have an α,β-unsaturated keto group, they can be inactivated by members of the Alpha, Mu and Pi class GSTs through conjugation with GSH [Atsmon *et al.*, 1990a; Atsmon *et al.*, 1990b; Parker & Ankel, 1992; Ohno & Hirata, 1993; Bogaards *et al.*, 1997]. The PGJ_2 series function as activating ligands for the nuclear hormone receptor peroxisome proliferator-activated receptor γ (PPARγ), which regulates adipocyte differentiation [Forman *et al.*, 1995; Kliewer *et al.*, 1995], hence the ability of the GSTs to conjugate the PGJ_2 series with GSH may block signalling to PPARγ.

The GSTs also have a more direct role in the regulation of signalling pathways. GSTP1-1 has been shown to regulate signalling of the stress kinase Jun N-terminal kinase (JNK). Existing as the GSTP1/JNK complex, monomeric GSTP1 acts as an inhibitor and thus protects cells against H_2O_2 induced death [Adler *et al.*, 1999]. In addition, GSTM1-1 has been shown to act as an inhibitor of Apoptosis Signal-regulating kinase 1 (ASK1) activity [Cho *et al.*, 2001]. By forming a complex with ASK1, GSTM1-1 regulates the downstream activation of the stress activated protein kinase p38. During heat shock, the ASK1/GSTM1-1 complex dissociates, allowing activation of p38 [Dorion *et al.*, 2002].

One of the more unusual non-detoxification roles has been attributed to the Omega class GSTs. Sequence similarities have been found between GSTO1-1 and the chloride intracellular channel 1 (CLIC1) nuclear channel. Further studies have revealed that whilst GSTO1-1 does not form ion channels, it inhibits ryanodine receptor activity in cardiac muscle (RyR2) and potentiates ryanodine receptor activity in skeletal muscle (RyR1). Hence it has been suggested that GSTO1-1 has a role in protecting cells from apoptosis induced by Ca^{2+} mobilisation from intracellular stores [Dulhunty *et al.*, 2001].

1.3.2.2.2.3 Ligand Binding

Some GSTs are able to noncovalently bind non-substrate ligands (neutral and anionic lipophilic chemicals) such as bilirubin, bile acids, hormones and penicillin. Although the precise function of ligand binding is not known, it has been suggested that such binding is important for solubilisation of water insoluble ligands to allow intracellular transport, sequestration of metabolic precursors for storage or sequestration of toxic molecules as a defense mechanism [Litwack *et al.*, 1971; Kamisaka *et al.*, 1975; Wolkoff *et al.*, 1979; Hayes *et al.*, 1980; Simons & Jagt, 1980; Homma & Listowsky, 1985; Listowsky *et al.*, 1988; Ishigaki *et al.*, 1989]. The Alpha class GSTs in particular are able to covalently bind a number of carcinogens such as 4-aminoazobenzene and 3-methylcholanthrene, in order to prevent these from causing DNA damage [Ketterer *et al.*, 1967; Jakoby, 1978]. These non-substrate ligands appear to bind to a unique site distinct from the active site and whilst they act as allosteric inhibitors of enzyme catalysis, substrate binding is not inhibited [Boyer *et al.*, 1984; Urade *et al.*, 1987]. Recent work on GSTs from various species proposes that there may be three ligand binding sites: one located in the intersubunit cleft [McTigue *et al.*, 1995; Ji *et al.*, 1996; Barycki & Colman, 1997; Le Trong *et al.*, 2002; Sayed *et al.*, 2002]; one located within the H-site [Oakley *et al.*, 1999]; and one known as the buffer binding site, located near the H-site between β2 and α1 [Ji *et al.*, 1997; Prade *et al.*, 1997; Oakley *et al.*, 1999].

1.3.2.3 HUMAN GST TISSUE DISTRIBUTION

GST activity is widespread throughout the human body, and has been detected in all human tissues tested (Table 1.1). The human GSTs are subject to differential expression and each tissue appears to have a characteristic GST profile in which the different isoenzymes have a unique cellular distribution and are expressed at different levels. For example, the liver contains high concentrations of GSTs, where they make up to 5%-10% of the total cytosolic protein content, consistent with its role as a major site of detoxification of xenobiotics and naturally occurring toxins [Whalen & Boyer, 1998; Hayes & Strange, 2000]. The distribution of the GSTs expressed in the human liver is further restricted with expression of GSTP1-1 limited to epithelial cells, GSTM1-1 to hepatocytes [Strange *et al.*, 1989], GSTT1-1 to bile duct epithelial cells and the nucleus of hepatocytes [Sherratt *et al.*, 2002] and GSTA4-4 to bile duct cells, vessels and periportal hepatocytes of the liver [Desmots *et al.*, 2001]. The cellular distribution of GST isoenzyme expression tends to reflect the role each plays in protecting against xenobiotics. For example, expression of GSTA4-4 in periportal cells as opposed to the centilobular cells is believed to be related to the high activity of GSTA4-4 with products of lipid peroxidation, as the periportal zone is implicated in the initial breakdown of fatty acids, has a higher rate of fatty acid oxidation and therefore generates high levels of ROS [Gebhardt, 1992; Desmots *et al.*, 2001].

Whilst patterns of GST expression tend to be conserved between individuals, they are subject to interindividual variation, as shown by studies on GST expression in the pancreas and the gastrointestinal tract (GI) [Coles *et al.*, 2000; Coles *et al.*, 2002]. The total GST content in the pancreas has been found to vary by seven-fold and the GST A1-1, GSTA2-2, GSTM2-2, GSTM3-3 and GSTP1-1 class isoenzymes have been found to vary six- to 30-fold between individuals. GSTA4-4 alone has been found to vary 50- to 100-fold from no expression through to very high expression in the pancreas [Coles *et al.*, 2000]. A high degree of interindividual variation and consistent organ-specific GST expression profiles have also been shown in the GI tract. Expression of GSTA1-1 in the stomach was found to vary two- to 160-fold between individuals. In addition, GSTA1-1 expression was consistently lower in the colon and stomach, with a 20- to 800 fold decrease in the colon compared to the small intestine and a 1.4- to 140-

Table 1.1 – Distribution of glutathione transferases in human adult tissues[a].

Tissue	Alpha				Mu					Omega		Pi	Sigma	Theta		Zeta	Unknown[b]
	A1-1	A2-2	A3-3	A4-4	M1-1	M2-2	M3-3	M4-4	M5-5	O1-1	O2-2	P1-1	S1-1	T1-1	T2-2	Z1-1	
Adipocytes												+	+				
Adrenal Gland	+	+	+	+	+	+	+					+					
Appendix													+				
Brain	+			+	+	+	+	+	+	+		+	+	+		+	
Cerebrum				+	+		+	+				+		+			
Bone Marrow				+									+				
Breast					+							+					
Cartilage													+				
Colon	+	+		+	+		+			+	+	+		+	+		
Erythrocytes												+		+			+ Basic
Gall bladder												+					+ Alpha
Heart/Aorta				+	+	+	+	+	+	+	+	+	+	+		+	
Kidney	+	+		+	+	+	+	+		+	+	+	+	+		+	
Lens												+					+ Basic
Liver	+	+		+	+		+	+		+	+		+	+	+	+	+ Alpha
Lung	+	+	+	+	+	+	+	+	+	+		+	+	+		+	+ Alpha, Mu
Lymphocytes					+							+					
Mammary Gland	+		+	+													
Macrophage													+				
Ovary			+					+		+	+	+					+ Alpha, Mu
Pancreas	+	+		+	+	+	+	+		+		+	+	+		+	+ Alpha
PBMC													+				
Pituitary gland	+				+	+	+					+	+				

Placenta	+			+	+				+		+			+	+		+	
Platelets															+			
Prostrate	+	+			+	+		+			+	+		+	+	+		+ Alpha
Retina															+			+ Basic
Salivary Gland				+								+						
Skeletal Muscle				+	+	+	+	+			+	+		+	+	+	+	
Skin	+														+			+ Alpha
Small Intestine	+	+		+	+	+	+				+			+	+	+		+ Basic
Spleen				+	+	+					+			+	+	+		
Stomach	+	+	+	+	+		+							+	+			+ Alpha
Synovium														+				
Testis	+	+	+	+	+	+	+	+			+	+		+				+ Alpha
Thymus				+		+					+			+	+			+ Alpha
Trachea	+	+	+	+								+		+				
Urinary bladder		+												+				+ Alpha
Uterus		+			+									+				+ Neutral

[a] Based on Table 2 from Johansson & Mannervik, 2001b. Other data was compiled from the following references: Warholm *et al.*, 1980; Laisney *et al.*, 1984; Stockman *et al.*, 1985; Suzuki *et al.*, 1987; Hayes *et al.*, 1989; Beckett *et al.*, 1990; Campbell *et al.*, 1990; Hussey & Hayes, 1992; Comstock *et al.*, 1993; Takahashi *et al.*, 1993; Tan *et al.*, 1995; Juronen *et al.*, 1996; Board *et al.*, 1997; Rowe *et al.*, 1997; Sherratt *et al.*, 1997; Desmots *et al.*, 1998; de Bruin *et al.*, 1999; Fernández-Cañón *et al.*, 1999; Board *et al.*, 2000; Coles *et al.*, 2000; Jowsey *et al.*, 2001; Coles *et al.*, 2002; Morel *et al.*, 2002.

[b] Cannot be assigned conclusively as these were found prior to identification of all currently known isoenzymes.

fold decrease in the stomach compared to the duodenum [Coles *et al.*, 2002]. Such interindividual differences can be linked to age, gender and physiopathological and genetic factors [Desmots *et al.*, 2001]. Induction of enzyme expression can also contribute to the observed differences between individuals, for example GSTP1-1 is induced by insulin but suppressed by retinoic acid [Xia *et al.*, 1993], whilst the Alpha class GSTs are induced by chemoprotective agents, such as the dithiolethione oltipraz [Morel *et al.*, 1993], and isothiocyanates found in brussel sprouts [Bogaards *et al.*, 1994]. Despite the highly variable nature of GST expression between humans, the ability to detoxify xenobiotics is maintained although individuals with lower expression levels of a particular GST will become more susceptible to the effects of the particular xenobiotic the poorly expressed GST metabolises. Hence the differential expression of GSTs has been suggested to influence individual and organ specific variation in susceptibility to xenobiotic induced disease [Hayes & Pulford, 1995; Hayes & Strange, 2000].

GST expression profiles and cellular distributions within a tissue vary during human development, with fetal tissues having a completely different GST profile when compared to their adult counterparts (Table 1.2) [Strange *et al.*, 1984; Strange *et al.*,

"Developmental changes of GST expression profiles can greatly vary between the different GST classes, with changes in some GST classes restricted to a single tissue, whilst changes in other GST classes occurring in multiple tissues [Strange *et al.*, 1985]."

expression of the Alpha class GSTs increases during gestation, expression of the Mu class increases from 30 weeks gestation and that of the Pi class decreases in the liver from 30 weeks gestation [Strange *et al.*, 1985]. During gestation the liver is a major site of haematopoiesis, therefore the observed changes in GST profiles may simply represent a change in cellular composition rather than gene expression [Hiley *et al.*, 1988]. GSTP1-1 is the most highly expressed and ubiquitous GST in both fetal and adult tissues [Laisney *et al.*, 1984; Cossar *et al.*, 1990; Di Ilio *et al.*, 1990]. Yet although it is the major GST in human fetal liver where it is expressed in the hepatocytes and bile duct epithelium, it is either absent or detectable at very low levels in the epithelium of adult liver [Awasthi *et al.*, 1980; Warholm *et al.*, 1981; Guthenberg *et al.*, 1986; Pacifici *et al.*, 1986; Faulder *et al.*, 1987; Hiley *et al.*, 1988; Strange *et al.*, 1989]. Expression of GSTS1-1, which is believed to play a role in haematopoiesis and immune function, also decreases from relatively high levels in fetal liver to almost

Table 1.2 – Distribution of the glutathione transferases in human fetal tissues[a].

Tissue	Alpha			Mu		Pi	Sigma	Zeta
	A1-1	A2-2	Unknown Alpha[b]	M1-1	Unknown Mu[b]	P1-1	S1-1	Z1-1
Adrenal Gland			+	+	+	+		
Bladder			+	+		+		
Bone Marrow							+	
Brain				+		+		
Brainstem			+	+		+		
Kidney	+	+	+	+	+	+		
Liver	+	+	+		+	+	+	+
Lung	+	+	+	+		+		
Heart			+	+		+		
Oesophagus			+	+		+		
Skeletal muscle								+
Small Intestine			+	+		+		
Spleen			+	+		+		
Stomach			+	+		+		
Testis				+				

[a] Based on Table 1 in Raijmakers *et al.*, 2002 and Table 3 in Johansson & Mannervik, 2001b. Other data was compiled from the following references: Warholm *et al.*, 1981; Guthenberg *et al.*, 1986; Faulder *et al.*, 1987; Strange *et al.*, 1989; Beckett *et al.*, 1990; Cossar *et al.*, 1990; Board *et al.*, 1997; Jowsey *et al.*, 2001.

[b] Cannot be assigned conclusively as these were found prior to identification of all currently known isoenzymes.

nothing in adult liver [Kanaoka *et al.*, 2000; Jowsey *et al.*, 2001]. In contrast, expression of GSTM1-1, which along with the Alpha class GSTs is the major GST found in adult liver [Rowe *et al.*, 1997], increases significantly after birth from relatively low levels [Warholm *et al.*, 1981; Strange *et al.*, 1989]. Expression of GSTA1-1 and GSTA2-2 does not change during gestation in fetal liver, but increases during the perinatal period. It has been suggested that an increase in expression of these enzymes during this period may enable the fetus to survive the transition into the aerobic post-natal environment [Strange *et al.*, 1989]. These developmental differences in GST expression imply that fetal tissues are sensitive to a different range of xenobiotics and therefore have very different detoxification requirements.

Gender related differences in GST expression levels have also have been observed in some human tissues but not in others. For example, significantly higher concentrations of total GST Alpha have been observed in female livers [Mulder *et al.*, 1999] and GST Alpha activity in the skin is 1.6-fold higher in females [Singhal *et al.*, 1993], yet in the colon, activity is two-fold higher in males [Singhal *et al.*, 1992b]. Also, higher plasma levels of GSTA1-1 have been observed in males, but the level in females was found to increase with age [Tiainen & Karhi, 1994; Mulder *et al.*, 1997]. In contrast, a Mu isoform of pI 6.2 has only been found in the female colon [Singhal *et al.*, 1992b]. Plasma samples of blood donors were also found to contain 1.5-fold higher levels of GSTP1-1 in males [Mulder *et al.*, 1997]. It has been suggested that some gender specific differences in GST expression may in fact be linked with gender specific differences in some cancers [Silverberg & Lubera, 1987].

1.3.2.4 GST Over-expression and Drug Resistance

The GSTs have been implicated in resistance of cells and organisms to drugs, pesticides, herbicides and antibiotics [Mannervik & Danielson, 1988; Piccolomini *et al.*, 1989; Fournier *et al.*, 1992; Dainelli *et al.*, 2002]. Human GSTs in particular have been implicated in the growing resistance and subsequent failure of some tumour cells to respond to chemotherapeutic drugs, which ultimately results in patient relapse. There is now substantial evidence that tumour cells often have an altered GST expression profile

compared to their normal counterparts and that over-expression of both GSH and GSTs is an important factor in the development of acquired anti-cancer drug resistance in preneoplastic lesions [Kitahara *et al.*, 1984] and tumour cells [Tew, 1994]. GSTP1-1 is the predominant GST found overexpressed in the majority of human tumours [Shea *et al.*, 1988; Sato, 1989; Castro *et al.*, 1990] and frequently displays an increase in the concentration (up to 3% of the total protein content) seen in normal tissue [Shea *et al.*, 1988; Lewis *et al.*, 1989; Kelley *et al.*, 1994] and is therefore used as a putative tumour marker. Over-expression of GSTP1-1 in a multidrug resistant human breast cancer cell line has been associated with a 45-fold increase in activity of GSTP1-1 [Batist *et al.*, 1986] and also increases the ability of cells to withstand insult by cytotoxic drugs [Nakagawa *et al.*, 1990]. However, due to the large interindividual variation of GST expression observed between tissues, the drug resistance phenotypes of tumour cells is dependent upon and differs according to the GST expression profile in the tissue of interest [Castro *et al.*, 1990; Tew, 1994]. Over-expression of the Alpha class GSTs is implicated in the development of tumour cell resistance to alkylating anti-neoplastic drugs, in particular the nitrogen mustard drugs such as chlorambucil and cyclophosphamide [Ciaccio *et al.*, 1991; Meyer *et al.*, 1992; Dirven *et al.*, 1994; Tew, 1994]. Usually, the Alpha class GSTs have a protective role in detoxifying these drugs. The Mu and Theta class GSTs have been implicated in resistance to the nitrosoureas. These GSTs, especially GSTT1-1 [Lien *et al.*, 2002] have a high activity with 1,3-bis(2-chloroethyl)-1 nitrosourea (BCNU). Over-expression of GSTM3-3 in human lung cells has been shown to catalyse the inactivation of this chemotherapeutic agent [Berhane *et al.*, 1993] and it is possible that the same may be true for GSTT1-1 and GSTM2-2 [Lien *et al.*, 2002]. As alkylating cancer drugs are also GST substrates, it is possible that the different side effects experienced by cancer patients is caused by variations in individual GST expression patterns [Johansson & Mannervik, 2001b]. By determining the GST isoenzyme profile of tumours, an appropriate treatment may be selected depending on the levels of expression and activity within the tumour [Kelley *et al.*, 1994]. GSTs are therefore a therapeutic target for rational drug design [Coles & Ketterer, 1990].

1.3.2.5 POLYMORPHISMS IN THE GSTS

There is great interest in whether the failure to express GSTs, or differences in GST expression, influences susceptibility to disease due to their ability to metabolise a diverse range of exogenous and endogenously produced xenobiotics. It is widely accepted that exposure to exogenous environmental chemicals in the form of carcinogenic PAHs and aromatic amines in tobacco smoke, automotive exhaust, industrial by-products and cooked meats, fish and baked or fired grain derived dishes [Felton *et al.*, 1994; Vineis, 1994] is associated with increased risk of cancer [Doll & Peto, 1981]. In addition, the accumulation of ROS in a wide range of tissues contributes to aging and many diseases associated with oxidative stress [Beckman & Ames, 1998]. The effects of genotoxic carcinogens can often be evaluated through measuring changes in the incidence of sister chromatid exchange, in the structure or number of chromosomes, micronuclei formation and altered enzyme activity [Norppa, 1997; Strange *et al.*, 2000]. Absent or dysfunctional members of the GST classes are therefore likely to alter individual tolerance to different xenobiotics and are candidates for increased disease susceptibility and altered drug responses.

The existence of polymorphism in the GSTs was first discovered when high GST Mu expression was observed in some, but not all, human livers. Electrophoretic analysis and activity assays identified this as the *GSTM1* null (*GSTM1*0*) allele [Warholm *et al.*, 1980; Board, 1981a; Strange *et al.*, 1984]. Subsequent polymorphisms have been identified through electrophoretic methods [Board, 1981a; Laisney *et al.*, 1984; Strange *et al.*, 1984], analysis of metabolic differences [Peter *et al.*, 1989; Bogaards *et al.*, 1993; Hallier *et al.*, 1993; Hallier *et al.*, 1994; McLellan *et al.*, 1997], differences in protein expression levels [Inskip *et al*, 1995; Wang *et al.*, 2000; Coles *et al.*, 2001a] and disease association studies [Latsoudis *et al.*, 2000]. Other polymorphisms have been discovered as sequence discrepancies between published sequences [Board *et al.*, 1989; Ali-Osman *et al.*, 1997; Coggan *et al.*, 1998; Johansson & Mannervik, 2001a]. More recently, analysis of the EST database has proved an effective tool for the discovery of polymorphism in the GSTs [Blackburn *et al.*, 2000; Blackburn *et al.*, 2001]. Coding region polymorphic variants in members of the Alpha, Mu, Pi, Theta and Zeta classes have been discovered and are listed in Table 1.3. Coding region polymorphisms in the

Table 1.3 – Known polymorphisms that affect the coding regions of the human GST genes.

Class	Gene	Nucleotide Altered[a]	Consequence[b]	References
Alpha	*GSTA?*	Gene deletion	Protein Deleted	Strange *et al.*, 1984
	GSTA1	c.365G>A	p.K125K	(Tetlow *et al.*, 2000); Bredschneider *et al.*, 2002
	GSTA2	c.328C>T	p.P110S	Wang *et al.*, 2000
		c.335G>C	p.S112T	Hayes *et al.*, 1989
		c.629A>C	p.E210A	Röhrdanz *et al.*, 1992
		*Hgi*AI site created	Unknown	Chen & Board, 1987
	GSTA3	c.1-87del	Exons 1 and 2 deleted	Johansson & Mannervik, 2001a
		c.88-87del	Exon 3 deleted	Board, 1998; Johansson & Mannervik, 2001a
		c.211C>A	p.L73I	Johansson & Mannervik, 2001a
	GSTA4	SNP in exon 1	Unknown	Latsoudis *et al.*, 2000
		c.487A>G	p.T163A	Iida *et al.*, 2001
Mu	*GSTM1*	Gene deletion	Protein deleted	Board, 1981a; Strange *et al.*, 1984; Seidegård *et al.*, 1988
		c.519G>C	p.K173N	Board, 1981a; Widersten *et al.*, 1991; Seidegård *et al.*, 1988
		Gene duplication[c]	Protein duplicated	McLellan *et al.*, 1997
	GSTM3	c.439T>G	p.G147W	Pearson *et al.*, 1993; Ross & Board, 1993
		c.670G>A	p.V224I	Emahazion *et al.*, 1999; Iida *et al.*, 2001
	GSTM4	c.178-259del	Exon 4 deleted	Ross & Board, 1993
		c.534T>C	p.F178F	Comstock *et al.*, 1993; Pearson *et al.*, 1993; Ross & Board, 1993; Zhong *et al.*, 1993a
		c.634G>A	p.V212V	Comstock *et al.*, 1993; Ross & Board, 1993; Zhong *et al.*, 1993a
		Ex8delinsVSCGIMX	Exon 8 deleted and replaced by 6 codons	Ross & Board, 1993

	GSTM5	*Hind*III site created	Unknown	De Jong *et al.*, 1991; Pearson *et al.*, 1993
Pi	*GSTP1*	c.313A>G	p.I105V	Ali-Osman *et al.*, 1997; Board *et al.*, 1989
		c.341C>T	p.A114V	Ali-Osman *et al.*, 1997; Board *et al.*, 1989
Theta	*GSTT1*	Gene deletion	Protein deleted	Pemble *et al.*, 1994
		c.310A>C	p.T104P	Alexandrie *et al.*, 2002
	GSTT2	c.417G>A	p.M139I	Coggan *et al.*, 1998
Zeta	*GSTZ1*	N.D.	p.K32K	Fernández-Cañón *et al.*, 1999
		c.94A>G	p.K32E	Board *et al.*, 1997; Fernández-Cañón *et al.*, 1999; Blackburn *et al.*, 2000
		c.124A>G	p.R42G	Board *et al.*, 1997; Fernández-Cañón *et al.*, 1999; Blackburn *et al.*, 2000
		c.245C>T	p.T82M	Fernández-Cañón *et al.*, 1999; Blackburn *et al.*, 2001

[a] The numbering of nucleotides changed in the variant alleles is based on their position in the cDNA from the A in the initiation codon.

[b] The numbering of amino acids includes the initiator methionine

[c] The duplicated gene possesses an additional copy of either the *GSTM1*A* or *GSTM1*B* genes, or both. When the duplicated allele is defined, the terms *GSTM1*1A* and *GSTM1*1B* should be used.

N.D.: Not Described.

Kappa, Omega and Sigma class GSTs have not yet been reported. Some of the polymorphisms detected to date have been associated with altering susceptibility to carcinogens and toxins, hence influencing the efficacy and toxicity of drug treatment and development of disease [Hayes & Strange, 2000]. However, this will be discussed in more detail in the relevant results chapters.

1.3.2.5.1 PHENOTYPE-GENOTYPE RELATIONSHIPS

Many studies have attempted to link various GST variants with disease susceptibility and outcome, and have provided evidence that allelism in these genes does in fact mediate disease phenotype [Strange *et al.*, 2001]. A number of groups have focussed on the effects of a single gene variant in small sample groups such as the *GSTM1*0* allele and have reported conflicting evidence of its association with lung and bladder cancers. However, whilst the results generated are statistically significant, the odds ratios are generally low (<2) and therefore provide weak evidence of genotype/phenotype associations [Rebbeck, 1997; Houlston, 1999; Hayes & Strange, 2000; Johns & Houlston, 2000; Benhamou *et al.*, 2002; Engel *et al.*, 2002]. In addition to investigating the single effects of GST polymorphisms, many studies have also extensively investigated the compound effects of multiple GST polymorphisms, particularly the *GSTM1* and *GSTT1* null alleles, and found that the level of association to disease susceptibility was dependent on interactions with other epistatic alleles or environmental factors [Brockmöller *et al.*, 1994; Hietanen *et al.*, 1997; Norppa, 1997; Rebbeck *et al.*, 1997; Hirvonen, 1999]. However, most studies have been unable to demonstrate a significantly stronger association to disease susceptibility caused by these interactive genetic effects above those caused by a single variation [Strange *et al.*, 2000].

Allelic variants in the GST genes, either singly or in combination with other genes, have been shown to have a greater impact on disease outcome in phenotypically homogenous subgroups of patients [Hayes & Strange, 2000; Strange *et al.*, 2001]. For example, cutaneous basal cell carcinoma patients cover a wide phenotypic spectrum but when patients were divided into subgroups dependent on further phenotypic classification, it was found that the frequency of the *GSTT1* null (*GSTT1*0*) allele and the *CYP2D6 EM*

polymorphism increased in patients displaying multiple cluster multiple presentation phenotype (MPP) as opposed to single cluster MPP and the odds ratios were correspondingly high (7.4 and 15.5 respectively), implying that these genes exert independent effects on susceptibility [Ramachandran *et al.*, 1999; Ramachandran *et al.*, 2001; Strange *et al.*, 2001].

Although many studies have investigated links between polymorphism in the xenobiotic metabolising enzymes with disease, the investigations have generally concentrated on one ethnic group rather than comparing the same association in several groups. However, wide ranging variations in allele frequency distribution among ethnic groups has also been observed, for instance the *GSTM1*0* allele is found at 50% in the Caucasian population [Board, 1981a; Strange *et al.*, 1984; Seidegård *et al.*, 1988], but ranges from approximately 21%-30 % in Chilean [Quiñones *et al.*, 1999] and some African groups [Zhao *et al.*, 1994; Mukanganyama *et al.*, 1997; Masimirembwa *et al.*, 1998; Johansson & Mannervik, 2001b] and 100% in the Micronesian population [Board *et al.*, 1990]. In many instances the associations between a particular genetic risk factor and cancer appears to be dependent on race or ethnicity. This effect has been extensively studied in the Phase I enzyme CYP1A1, which has five polymorphisms, one of which, *CYP1A1*4*, is specific to African individuals [Crofts *et al.*, 1993]. This African-American specific polymorphism has been associated with increased risk of lung and breast cancer [Taioli *et al.*, 1995a; Taioli *et al.*, 1995b]. Another CYP1A1 polymorphism, *CYP1A1*2*, has also only been associated with lung cancer in Japanese patients [Kawajiri *et al.*, 1990], but has not been found by other studies in the Finnish [Hirvonen *et al.*, 1992] and Norweigan populations [Tefre *et al.*, 1991]. Such inter-racial differences may possibly be due to differences in linkage disequilibrium patterns, gene regulation and function, and possibly the fact that different alleles may have diverse effects within different populations [Garte, 1998]. Thus within the human population, allele frequencies are generally influenced by ethnicity, race and geography [Garte, 1998].

1.3.3 Detection of SNPs in the GSTs

The GST superfamily provides a highly efficient line of defence against the toxic insults to which we are daily exposed. Polymorphisms within these proteins can greatly alter the efficiency of this defence therefore detection of novel SNPs in the GST genes is likely to help elucidate the mechanisms responsible for the observed variation in drug response and altered susceptibility to a number of diseases. This can now be easily achieved by utilising the new advances in SNP detection. Although a number of polymorphisms have previously been reported in human GST genes, they have generally been identified by chance. This study was aimed at applying a bioinformatic approach to identify the extent of genetic polymorphism in the human GST gene family. A large number of polymorphisms in the GST genes were detected in the current study, allele frequencies were determined in four ethnic populations and the functional effects of these polymorphisms were characterised.

CHAPTER 2

MATERIALS & METHODS

2.1 COMMONLY USED MATERIALS

This chapter describes the materials and methods routinely used during this research. Specific procedures and resources will be described in the relevant chapters. Commonly used chemicals compounds are listed in Appendix A1, and are of analytical grade. Enzymes are listed in Appendix A2; buffers, solutions and culture media are listed in Appendix A3.

2.2 LABORATORY PROCEDURES

Solutions requiring sterilisation were either autoclaved at 121°C/100 kPa for 15 minutes, or filter sterilised through 0.2 µM membranes. Disposable plastic labware was autoclaved before use and glassware was baked overnight at 180°C. Distilled deionised water (ddH_2O) was used for all solutions. All experiments were carried out under PC2 laboratory conditions as stipulated by the Genetic Manipulation Committee. Radioactive substances were handled in accordance with the Australian National University Radiation Safety Handbook.

2.3 DATABASE ANALYSIS

2.3.1 SEQUENCE DATABASE ANALYSIS PROGRAMS

Potential polymorphisms were identified in the human EST database using the default settings of the advanced BLAST alignment tool [Altschul *et al.*, 1997], which can be accessed through the NCBI web site http://www.ncbi.nlm.nih.gov/. Aligned EST sequences were examined for nucleotide variations using the "flat query-anchored with identities" format. Potential polymorphisms identified using this method were verified by ordering and sequencing I.M.A.G.E. (Integrated Molecular Analysis of Genomes and their Expression) Consortium (LLNL; Lawrence Livermore National Library) cDNA Clones [Lennon *et al.*, 1996] (§2.5.5.1) through Genome Systems, Inc./Incyte Genomics

(Palo Alto, USA). The SNP Finder program, available at http://lpg.nci.nih.gov/GAI/, was also used to identify potential variants in the GSTs using UniGene clusters that represented the gene of interest [Buetow *et al.*, 1999]. This program provides access to electropherograms hence potential polymorphisms could be verified through examination of existing trace data.

2.3.2 SNP DATABASES

Various SNP databases containing compilations of polymorphisms collated by SNP discovery programs were also searched for GST polymorphism entries by entering the gene name of interest. These included dbSNP, which was accessed through the LocusLink site at the URL http://www.ncbi.nlm.nih.gov/LocusLink/; the CGAP-GAI candidate (CGAPC), validated (CGAPV) and confirmed (CGAPCo) SNP lists, available at http://lpg.nci.nih.gov/; HGBASE (now HGVbase), which was accessed at http://hgvbase.cgb.ki.se/; refseq, available at http://lpgws.nci.nih.gov:82/perl/snp2ref; the UUGC GeneSNPs database, at the URL http://www.genome.utah.edu/genesnps/; and the UWGC EGPSNPs database, which is available at the URL http://www.genome.washington.edu/projects/egpsnps/. The SNPper program, at http://bio.chip.org/biotools, was also utilised to detect potential polymorphisms. Polymorphisms identified through the CGAP-GAI lists could be confirmed with existing trace data, those identified in the remaining lists were confirmed by small-scale population analysis (§2.5.6.2).

2.4 BACTERIAL STRAINS AND PLASMIDS

Escherichia coli strains DH5α, M15(pRep4), TG1 and XL1 Blue (Table 2.1) were used to host recombinant proteins in the vectors pRB269, pKK261, pKK263 and pQE30 (Qiagen). These were grown on LB plates and cultured in LB media overnight at 37°C. Media was supplemented with 100 μg/mL ampicillin, 34 μg/mL chloramphenicol or 20 μg/mL kanamycin as required to select for cells containing plasmids. Important clones were stored at -70°C in 50% (v/v) glycerol. Individual colonies were isolated by

dilution streaking from bacterial stabs or glycerol stocks onto LB-agar plates supplemented with the appropriate antibiotic and an overnight incubation at 37°C.

Table 2.1 – Strains of *E. coli* used for recombinant protein expression.

Strain	**Genotype**
DH5α	ϕ80d*lacZ*ΔM15, *recA*1, *endA*1, *gyrA*96, *thi*-1, *hsdR*17(r_K^-, m_K^+), *supE*44, *relA*1, *deoR*, Δ(*lacZYA-argF*)U169
M15(pRep4)	NaI^S, Str^S, rif^S, lac^-, ara^-, gal^-, mtl^-, F^-, $recA^+$, uvr^+
TG1	F', *tra*D36, *lacI*^{q}ZΔM15, *pro*AB$^+$/*sup*E, Δ(*hsd*M-*mcr*B)5, (r_K^-, m_K^+, *mcr*B$^-$), *thi*, Δ (*lac-pro*AB)
XL1 Blue	*rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17(r_K^-, m_K^+), *sup*E44, *rel*A1, *lac*, [F', *pro*AB$^+$, *lacI*^{q}ZΔM15, :Tn10(Tet^r)]

The ubiquitin fusion vector pRB269 was used to express the GSTA2-2 recombinant proteins [Baker *et al.*, 1994], which were constructed from the GSTA2D variant created by Dr Dan Liu. The GSTM3-3 recombinant proteins were expressed in pKK261, and were constructed from the GSTM3A variant created by Dr Veronica Ross. pKK261 is a modification of the pKK223-3 vector generated by removing a *Bam*HI-*Sal*I fragment from the backbone of pKK223-3. The GSTA3-3 protein was cloned into pKK263, which is a modified version of pKK261. The GSTO1-1 recombinant proteins were expressed in pQE30 (Qiagen), and were constructed from the GSTO1*A allelic isoform created by Professor Philip Board.

2.5 DNA MANIPULATIONS

2.5.1 DNA PREPARATION

2.5.1.1 PREPARATION OF LAMBDA DNA

cDNA from previously prepared lambda phage lysate originating from an LE392:λgt11 human testis cDNA library (Clontech, Palo Alto, CA, USA) was extracted using the Wizard™ Lambda Preps DNA Purification System in accordance with the manufacturer's protocols.

2.5.1.2 PREPARATION OF PLASMID DNA

Small-scale preparations of plasmid DNA was isolated from a 5 mL overnight culture of *E. coli* cells using the Promega Wizard Miniprep DNA Purification Kit, the GeneWorks Bresaspin™ Plasmid Minikit or the UltraClean™ Mini Plasmid Prep Kit (MoBio Laboratories Inc, USA). These kits were used in accordance with the manufacturer's protocols, although one addition was made when preparing DNA from plasmids propagated in TG1 cells – 5 µL of alkaline protease solution (275 µg/µL) was added between the lysis and the neutralisation steps. Alkaline phosphatase non-specifically degrades proteins and is known to inactivate the endonuclease A present in TG1 cells. After a five minute incubation at ambient temperature, the extraction process continued according to the manufacturer's protocols.

2.5.2 ELECTROPHORESIS

2.5.2.1 AGAROSE GEL ELECTROPHORESIS

Polymerase Chain Reaction (PCR) products and plasmid DNA were routinely separated on 0.8%-1.5% agarose gels in 10x TAE running buffer at a constant voltage of 70 V (or

10 V/cm of the distance between the electrodes). 1 Kb, 1 Kb plus (1µg/µL) and pUC19/*Hpa*II (500 ng/µL) were used as DNA size markers. Gels were stained with ethidium bromide (5 µg/ml was added to the running buffer) and visualised under UV illumination (254 nm). DNA fragments excised from agarose gels were extracted using the Geneclean® II kit (BIO 101, USA), the Qiagen Gel Extraction Kit or the UltraClean™ GelSpin DNA Purification Kit (MoBio Laboratories Inc, USA).

2.5.2.2 Acrylamide Gel Electrophoresis

Digested PCR product fragments for RFLP analysis were separated by 6%-12% polyacrylamide gel electrophoresis (PAGE) against the molecular size marker pUC19/*Hpa*II using a Hoefer SE600 series Vertical Slab Gel Unit (Hoefer Scientific Instrument, San Francisco, CA; 160 x 180 x 1.5 mm). Electrophoresis was performed at room temperature in 1x TBE running buffer at 30 mAmp for approximately 2 hours, or until the bromophenol blue dye front reached the bottom of the gel. Fractionated DNA was photographed using polaroid film under UV illumination after staining with ethidium bromide.

2.5.2.3 Recombinant Protein Expression Gels

Recombinant proteins were identified by separation on 12.5% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on either a PHAST System Separation and Control Unit (Pharmacia) or an Xcell II™ Mini-Cell apparatus (Novel Experimental Technology, USA), against the Rainbow™ coloured protein molecular weight marker. Gels were stained with Coomassie Brilliant Blue. GST Alpha and Mu recombinant protein concentrations were determined using the Peterson Assay [Peterson, 1977]. GST Omega recombinant protein concentrations were determined using the Lowry Assay [Lowry *et al.*, 1951].

2.5.3 DNA QUANTITATION

Oligonucleotide concentrations were determined by measuring the optical density (OD) of the DNA at an absorbance of 260 nm using a Varian Cary 1 UV-Visible Spectrophotometer. Concentrations were calculated in pmol/µL based on the absorbance measurements, given that 1 OD_{260} unit for single stranded DNA was equal to 33 µg/mL.

Concentrations of plasmid DNA were determined by comparison to the molecular weight marker λ DNA/*Hind*III (500 µg/mL), after electrophoresis through a 0.8%-1.5 % agarose gel in 1x TAE running buffer.

2.5.4 RESTRICTION ENDONUCLEASE DIGESTIONS

2.5.4.1 SUBCLONING

Plasmid DNA (approximately 1 µg) was routinely digested with 10-20 units of the appropriate restriction endonuclease following the manufacturer's protocol. Double digests were either performed simultaneously in a buffer compatible with both enzymes, or as subsequent reactions. Reactions were terminated by heat inactivation (65˚C for 20 minutes unless otherwise specified), phenol/chloroform extraction and ethanol precipitation, or by addition of a sucrose gel-loading dye.

2.5.4.2 RFLP ASSAYS

Genomic PCR products were tested for the presence of polymorphisms using RFLP assays. PCR products (2-5 µL) were digested in a 10 µL volume with 0.5-1 units of the appropriate restriction enzyme using the buffers and conditions as recommended by the manufacturer. Double digests were performed by selecting a buffer compatible with both enzymes. The reaction was terminated by addition of ficoll gel-loading dye.

2.5.5 DNA SEQUENCING

2.5.5.1 CYCLE SEQUENCING

I.M.A.G.E. Consortium (LLNL) cDNA Clones and GSTA2-2, GSTM3-3, GSTO1-1 and GSTO2-2 recombinant protein constructs were sequenced from double stranded DNA using the Thermosequenase cycle sequencing kit (Amersham) with [α-^{33}P] dATP. Sequence information was resolved on 6% polyacrylamide glycerol tolerant gels containing 8.3 M urea using an IBI Base Runner™ Nucleic Acid Sequencer at a constant of 35 W in 0.8x glycerol tolerant running buffer. Gels were fixed in a 10% acetic acid (v/v)/10% methanol (v/v) solution, before being vacuum dried on a Model 583 gel dryer (Bio-Rad Laboratories, USA) at 80°C for 1 hour. Autoradiography was performed at room temperature for at least 24 hours on Fuji Medical X-ray film (RX).

2.5.5.2 AUTOMATED SEQUENCING

GSTA3-3 and GSTO2-2 recombinant protein constructs were sequenced from double stranded DNA using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit using ABI Big Dye Terminator v.3 (Applied Biosystems, USA) and an ABI 377 DNA Sequencer at the Biomolecular Resource Facility at the John Curtin School of Medical Research.

2.5.6 PCR CONDITIONS

2.5.6.1 PREPARATION OF PRIMERS

Oligonucleotides were synthesised on an Applied Biosystems 380B Oligonucleotide Synthesiser by the Biomolecular Resource Facility at the John Curtin School of Medical Research and were quantitated as described in §2.5.3. Alternatively, primers were purchased from Geneset Pacific PTY. Ltd (Lismore, Australia). Primers were stored at -20°C and diluted as required. Annealing temperatures (T_m) for each primer were

calculated using the following equation, which is based on the nucleotide composition of the primer:

$$T_m = \{[(\#dCTPs + \#dGTPs) \times 4^{\circ}C] + [(\#dATPs + \#dTTPs) \times 2^{\circ}C]\} - 5^{\circ}C$$

where # refers to the number of nucleotides

2.5.6.2 AMPLIFICATION OF DNA

Amplification of DNA by PCR was conducted using a GeneWorks DNA Engine, a DNA Engine Tetrad (MJ Research PTC-225) or a Corbett Research FTS-1 thermal cycler.

Exons of the various GST genes were specifically amplified from genomic DNA obtained from individuals with representing the three major global populations – Africans, Chinese and Caucasians. The four populations chosen were: the Bantu Africans; the Madagascan Creole Africans – predominantly descended from the Bantu Africans but also an admixture of Caucasian and Asian/Oceania genetic influences [Hewitt *et al.*, 1996]; the Australian Europeans; and the Southern Chinese. 25 samples from the Australian European, Creole African and Southern Chinese ethnic groups were amplified by PCR and genotyped by RFLP assays for small-scale population screens. If a variant was detected in this small sample set, up to 100 individuals from each of the three ethnic groups were amplified and genotyped. 74 samples from the Bantu African population were genotyped at a later stage. Statistical comparisons between the four ethnic groups were performed using Chi-squared analysis. Results were defined as being statistically significant if $p<0.05$.

In general, a 20 µL reaction mixture contained 6 pmol of the forward and reverse primers specific for each exon, 25 ng template DNA, 1x *Taq* DNA polymerase reaction buffer, 1.5-2 mM $MgCl_2$ (25 mM), 0.2 µM dNTPs, 0.5 units of *Taq* DNA polymerase and ddH_2O. A negative control sample containing all components of the reaction except the template was included in each PCR to ensure the absence of contamination. PCR conditions involved an initial denaturation step at 95°C for 2 minutes followed by 35-40 cycles of a 95°C denaturation step (20 sec), 48-63°C annealing step (18-20 sec),

and 72°C extension step (30 sec), with a final 3 minute extension at 72°C. Agarose gel electrophoresis was used to confirm the absence of spurious bands and contamination, and the presence of PCR product (5 μL).

Nested PCR reactions were performed using the fragment amplified during the first reaction as the template and two new PCR primers located within the first primer pair. Reactions were conducted as outlined above and allowed greater yields of PCR product to be obtained. Specific details for amplification of GST exons can be found in the relevant results chapters.

2.6 Recombinant DNA Methods

2.6.1 Ligation

Vectors and inserts were prepared for ligation by digestion with the appropriate restriction endonuclease as described in §2.5.4.1. The digested insert DNA was isolated by agarose electrophoresis and the excised fragments purified using either the Geneclean® II kit (BIO 101, USA) or the UltraClean™ GelSpin DNA Purification Kit (MoBio Laboratories Inc, USA). Digested vector DNA was purified either as above or by standard phenol/chloroform extraction and ethanol precipitation. Vectors were resuspended in ddH_2O and stored at -20°C. Inserts were ligated into 100 ng vector at 4°C overnight. A 10 μL ligation reaction contained either 2 μL 5x BRL buffer and 0.5 μL of T4 DNA ligase (7500 Weiss units/mL) or 1 μL Ligase 10x buffer and 1 μL T4 DNA ligase (3 Weiss units/μL).

2.6.2 Transformation

Competent *E. coli* cells were prepared using the calcium, potassium and manganese chloride method. Aliquots were snap frozen on dry ice before storage at -70°C. Ligations (1μL) were transformed into 100 μL of either cell type and incubated on ice for 30 minutes. Samples were heatshocked at 45°C for 42 seconds, cooled on ice then

incubated with 500 µL of LB at 37°C for 45 minutes. Transformed cells were plated onto LB-Amp, LB-Amp/Chlor or LB-Amp/Kan plates as required and incubated overnight at 37°C.

2.6.3 Site-Directed Mutagenesis

All polymorphic variants causing amino acid changes were created using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). A 50 µL reaction contained 5 µL of 10x Reaction Buffer, 1.25 µL of the forward and reverse primers (10 pmol/µL), 2 µl of 10 mM dNTP mix, 5-10 ng dsDNA template and 1 µL *Pfu Turbo* DNA polymerase. Cycling parameters were adjusted depending on the type of mutation required. Introduction of a point mutation involved an initial denaturation step of 95°C (30 sec), followed by 12 cycles of 95°C for 30 seconds, 55°C for 1 minute and 68°C for 11 minutes. Introduction of a deletion involved an initial denaturation step of 95°C (30 sec), followed by 18 cycles of 95°C for 30 seconds, 55°C for 1 minute and 68°C for 11 minutes. Products were digested with *Dpn*I and transformed into TG1, XL1-Blue or M15Rep4 *E. coli* cells. All mutations and constructs were confirmed by DNA sequencing as described in §2.5.5.

2.6.4 Enzyme Characterisation

The specific activity of all recombinant proteins towards a range of substrates was determined spectrophotometrically at 37°C using previously described methods. These are described in Table 2.2. Statistical comparisons were performed using unpaired Student's t-test, and results were defined as being statistically significant if $p<0.05$.

2.6.5 Structure Analysis

The structural effects of verified polymorphisms were determined by predictive modelling using the DeepView/Swiss-PdbViewer program (GlaxoSmithKline), which

Table 2.2 – Substrates used for enzymatic characterisation of recombinant GST proteins.

Substrate	Final [substrate] (mM)	Final [GSH] (mM)	Buffer	A_{nm}	$\Delta\varepsilon$ ($mM^{-1}cm^{-1}$)	Reference
Δ^5-Androsten-3,17-dione	0.1	1	0.025 M NaPi pH8.0	248	16.3	Johansson & Mannervik, 2001a
1-chloro-2,4-dinitrobenzene	1	1	0.1 M NaPi pH6.5	340	9.6	Habig & Jakoby, 1981
7-chloro-4-nitrobenzo-2-oxa-1,3-diaxole	0.2	0.5	0.1 M NaOAc pH5.0	419	14.5	Ricci *et al.*, 1994
Cumene hydroperoxide	1.5	1	0.1 M NaPi pH7.0[a]	340	6.2	Beutler, 1975
1,2-dichloro-nitrobenzene	1	5	0.1 M NaPi pH8.0	344	10	Habig & Jakoby, 1981
1,2-epoxy-3-(*p*-nitrophenoxy)propane	0.5	5	0.1 M NaPi pH6.5	360	0.5	Habig & Jakoby, 1981
Ethacrynic acid	0.2	0.25	0.1 M NaPi pH6.5	270	5	Habig & Jakoby, 1981
2-hydroxyethyl disulfide	0.75	1	0.1 M Tris pH8, 2 mM EDTA[b]	340	6.2	Holmgren & Aslund, 1995
p-nitrophenylacetate	0.2	0.1	0.1 M NaPi pH7.0	400	8.79	Habig & Jakoby, 1981
4-phenyl-but-3-en-2-one	0.05	0.25	0.1 M NaPi pH6.5	290	-24.8	Habig & Jakoby, 1981
tert-butyl-hydroperoxide	1.5	1	0.1 M NaPi pH7.0*	340	6.2	Beutler, 1975
trans-non-2-enal	0.025	1	0.1 M NaPi pH6.5	225	-19.2	Brophy *et al.*, 1989

[a] add 0.3 U/mL glutathione reductase and 0.2 mM NADPH

[b] add 13.6 μg/mL glutathione reductase, 0.3 mM NADPH and 0.1 mg/mL BSA. This reaction was measured at 30°C.

can be accessed at http://www.expasy.org/spdbv/. Protein Data Bank (PDB) files for the GSTA1 (PDB code 1GSE), GSTM3 (PDB code 3GTU) and GSTO1 (PDB code 1EEM) proteins were downloaded from the PDB (http://www.rcsb.org/pdb). To generate a model for GSTA2, the 11 amino acids that differentiate GSTA2 from GSTA1 were introduced into the GSTA1 model, and the new protein sequence was submitted to the First Approach Mode at the Swiss-Model website (http://www.expasy.org/swissmod/) [Guex & Peitsch, 1997]. Similarly, in order to generate a model for each of the variant proteins, the relevant residue was changed in the original protein sequence, and the newly altered sequence was submitted to the First Approach Mode at the Swiss-Model website.

2.7 NOMENCLATURE

Genes, mRNA, cDNA, proteins and polymorphisms were all described according to the agreed nomenclature guidelines [den Dunnen & Antonarakis, 2001; Wain *et al.*, 2002].

2.8 COMPUTER SOFTWARE ANALYSIS

Various computer software programs were used in this study. DNA sequence translations and restriction maps were generated using DNA Strider version 1.2 (Commissariat a l'Energie Atomique, France). Oligonucleotides were designed using Amplify 1.2 (University of Wisconsin, USA) [Engels, 1993]. MacCurve Fit version 1.1.2 (Kevin Raner Software, Australia) was used to fit enzyme kinetic data. Student's t-tests were calculated using GraphPad Prism v. 3.02.

CHAPTER 3

THE ALPHA CLASS GSTS

3.1 INTRODUCTION

The Alpha class GSTs can be described as one of the more versatile GST families. With four expressed isoforms, this family is able to fulfill a variety of functions and has a correspondingly wide range of substrate specificities. Formerly known as ligandins, some Alpha class GSTs are able to bind non-substrate compounds such as bilirubin, bile acids, heme, thyroid and steroid hormones and penicillin in a non-covalent fashion, thus allowing solubilisation and storage of these compounds in the liver [Litwack *et al.*, 1971; Wolkoff *et al.*, 1979; Hayes *et al.*, 1980; Simons & Jagt, 1980; Homma & Listowsky, 1985; Ishigaki *et al.*, 1989]. The ligandins are also capable of covalently binding strong alkylating agents, providing protection against their damaging effects [Ketterer *et al.*, 1967; Jakoby, 1978]. Alpha class GSTs have been shown to catalyse the detoxification of the nitrogen mustard group of alkylating anticancer drugs [Ciaccio *et al.*, 1991; Meyer *et al.*, 1992; Dirven *et al.*, 1994; Tew, 1994], α,β-unsaturated aldehydes [Board, 1998; Hubatsch *et al.*, 1998] and some heterocyclic amines (HCAs) [Lin *et al.*, 1994; Coles *et al.*, 2001b]. In addition to their ability to function as typical GSTs, they all possess some degree of keto-steroid isomerase activity [Benson *et al.*, 1977; Johansson & Mannervik, 2001a] and prostaglandin synthase activity [Burgess *et al.*, 1989]. However, it is their selenium-independent peroxidase (GPx-II) activity that distinguishes the Alpha class GSTs from the other GST families [Awasthi *et al.*, 1980; Zhao *et al.*, 1999]. Given the diverse nature of the Alpha class GSTs, polymorphism in these genes may have far-reaching implications, both in terms of drug metabolism and disease aetiology.

The existence of polymorphism in the Alpha class GSTs has been previously recognised, notably in the *GSTA2* gene, which is predominantly expressed in the liver [Rowe *et al.*, 1997] and subject to interindividual variation in its level of expression [Hayes *et al.*, 1989]. Considering the Alpha class GSTs are believed to represent a major line of defence against oxidative stress induced by phospholipid hydroperoxides and fatty acid hydroperoxides, *GSTA2* polymorphism studies may prove to be of clinical interest. Through its GPx-II activity, *GSTA2*, and to a lesser extent *GSTA1* [Stockman *et al.*, 1987; Chow *et al.*, 1988; Zhao *et al.*, 1999], can catalyse the reduction of these organic hydroperoxides to the corresponding alcohols [Awasthi *et al.*, 1980; Singhal *et*

al., 1992a]. This activity is of particular importance to the liver, which being the major site for drug and xenobiotic detoxification, is at an increased risk of suffering oxidative stress due to the large number of ROS generated during detoxification and lipid peroxidation [Zhao *et al.*, 1999]. The first *GSTA2* polymorphism was described in 1989 [Hayes *et al.*, 1989], and differed from the original cDNA sequence at residue 112 [Rhoads *et al.*, 1987], with a serine replacing the threonine at this position. A later study, in which the full *GSTA2* gene sequence was determined, revealed a second polymorphism at residue 210, with an alanine replacing glutamic acid [Röhrdanz *et al.*, 1992]. A subsequent study found that the p.T112 variant was never seen in combination with the p.A210 variant, yet the p.S112 variant was associated with either the p.E210 or p.A210 variants [Coles *et al.*, 2000]. More recently, a polymorphism in which a serine replaces a proline at residue 110 was described [Wang *et al.*, 2000]. Whilst there is speculation as to the functional consequences of these three polymorphisms, there has been no further investigation into any of these variants. Alterations in the GPx-II activity of GSTA2-2 introduced by polymorphisms could ultimately be responsible for diseases associated with oxidative damage.

Another important component in the defence against oxidative stress is provided by the ubiquitously expressed GSTA4-4. This isoform has a uniquely high catalytic efficiency for the 4-hydroxyalkenals, particularly 4HNE [Board, 1998; Hubatsch *et al.*, 1998], one of the major byproducts of lipid peroxidation [Esterbauer *et al.*, 1991], and it has been suggested that this enzyme provides significant protection against oxidative damage [Desmots *et al.*, 1998; Hubatsch *et al.*, 1998]. 4HNE is a genotoxic α,β-unsaturated aldehyde and has been implicated in oxidative damage associated with aging [Stadtman, 1992], atherosclerosis [Palinski *et al.*, 1989], cataract formation [Ansari *et al.*, 1996], cancer [Eckl *et al.*, 1993] and neurodegenerative disease such as Parkinson's Disease [Yoritaka *et al.*, 1996] and Alzheimer's Disease [Mark *et al.*, 1997]. Recent localisation of GSTA4-4 to the mitochondria suggested that GSTA4-4 might help maintain mitochondrial redox status during cellular respiration, with disruption potentially resulting in oxidative stress [Gardner & Gallagher, 2001]. In addition, it has been suggested that the level of GSTA4-4 expression and any polymorphic variants of this gene may act as susceptibility factors towards degenerative disorders and diseases characterised by a loss of mitochondrial function [Gardner & Gallagher, 2001].

Recently, two coding region polymorphisms have been identified in the *GSTA4* gene. A preliminary report has described the presence of an uncharacterised SNP in exon 1 [Latsoudis *et al.*, 2000], and a threonine to alanine substitution at residue 163 (c.487A>G) was identified through a large-scale genome screening effort [Iida *et al.*, 2001], yet remains uncharacterised.

Polymorphisms in the Alpha class GSTs are also implicated in diseases other than those associated with oxidative stress. Recently, polymorphism in the promoter region of *GSTA1*, consisting of two alleles *GSTA1*A* and *GSTA1*B*, has been shown to not only affect both the absolute levels of GSTA1-1 expression and the ratio of GSTA1-1 to GSTA2-2 expression in the liver [Coles *et al.*, 2001a] and potentially other tissues [Morel *et al.*, 2002], but has also been linked to an increased susceptibility to colorectal cancer (CRC) [Coles *et al.*, 2001b]. CRC susceptibility has been associated with a high dietary intake of cooked meats, which are laden with heterocyclic amines (HCA) and their activated products, particularly N-acetoxy-2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (N-acetoxy PhIP) [Layton *et al.*, 1995], implying that HCAs may be involved in CRC aetiology. It is believed that low, or altered, expression of GSTA1-1 may influence its unique ability to catalyse N-acetoxy PhIP, thereby increasing the risk of CRC [Lin *et al.*, 1994; Coles *et al.*, 2001b]. Another three promoter region polymorphisms have since been identified, however these do not appear to have any significant effects on GSTA1-1 function [Bredschneider *et al.*, 2002]. Although polymorphisms have not yet been identified in the coding region of *GSTA1*, it can be postulated that these too could alter an individual's susceptibility to CRC.

In addition to the role GSTs play in detoxification, some have been shown to have a purely biological role. GSTA3-3 has a uniquely high activity as an isomerase in steroid hormone biosynthesis. GSTA3-3 specifically catalyses the double bond isomerisation of Δ^5-androstene-3, 17-dione and Δ^5-pregnane-3, 20-dione to Δ^4-androstene-3, 17-dione and Δ^4-pregnane-3, 20-dione, which are immediate precursors of testosterone and progesterone, respectively [Johansson & Mannervik, 2001a]. During the recent cloning of the full-length *GSTA3* cDNA, two novel *GSTA3* alternative transcripts were also identified along with a possible p.L71I polymorphism [Johansson & Mannervik,

2001a]. One transcript is characterised by the deletion of exon 3 which would, if translated, result in a +1 frameshift and give rise to a truncated protein of 32 amino acids [Johansson & Mannervik, 2001a]. In the second transcript both exons 1 and 2 are missing and 26 nonsensical nucleotides are attached to the 5' end of exon 3 [Board, 1998; Johansson & Mannervik, 2001a]. All three GSTA3-3 transcripts are expressed in steroidogenic tissues: the testis, ovary, adrenal gland and placenta, so it could be expected the alternative transcripts and any polymorphic variants may alter expression levels and protein function and hence the role of GSTA3-3 in steroidogenesis.

This chapter describes the use of various database mining techniques to facilitate the discovery of novel Alpha class polymorphisms. The distribution of the confirmed polymorphisms was determined in four ethnic populations and the recombinant proteins were characterised enzymatically and by predictive modelling to determine whether the function or the structure of the protein was altered by the introduction of the polymorphism.

3.2 MATERIALS AND METHODS

Database analysis and standard PCR and recombinant methods used in this chapter are described in Chapter 2. Techniques unique to this chapter are outlined below.

3.2.1 DATABASE SCREENING

Potential polymorphisms in the four Alpha class GSTs were identified and verified using the database mining methods described in §2.3, and are listed in Tables 3.1 and 3.2.

3.2.2 PCR/RFLP ANALYSIS

Potentially polymorphic exons of the *GSTA1*, *GSTA2* and *GSTA3* genes were specifically amplified from genomic DNA using PCR as described in §2.5.6.2. For variants detected in the *GSTA1* and *GSTA2* genes using the BLAST alignment tool and the SNP Finder program, 200 Australian Europeans, and up to 100 Creole Africans and Southern Chinese samples were amplified. Small-scale population screens, as described in §2.5.6.2, were performed to confirm the presence of variants detected in the SNP databases. The oligonucleotide sequences and the specific conditions for each PCR are listed in Table 3.3. The presence of variation in the PCR products was determined by RFLP analysis (§2.5.4.2) using the appropriate restriction enzymes (Table 3.4). Eight of the 15 verified sequence alterations changed restriction endonuclease sites. For the remaining seven, primers were designed containing a partial restriction endonuclease site, which was completed by either the absence or the presence of the polymorphic base [Kangsadalampai *et al.*, 1998]. Up to 200 samples from the Australian European, and 100 samples from the Bantu African, Creole African and Southern Chinese groups were amplified by PCR, to allow allele and haplotype frequency determination for any polymorphisms identified.

Table 3.1 – GST Alpha variants detected in the EST database by the BLAST alignment tool and in the UniGene database by the SNP Finder program.

Gene	Nucleotide	Residue	Exon	Program	ESTs with Alteration	Libraries represented	Sequence Confirmed[a,b]	Confirmed in population studies
GSTA1[c]	c.115G>T	p.E39X	3	BLAST	3	1	Yes-H73011	No
				SNP Finder	4	1	Yes-1.0	No
	c.159G>T	p.Q53H	4	SNP Finder	1	1	Yes-0.83	No
	c.365G>A	p.K125K	5	BLAST	29	7	Yes-T81799	Yes
	c.365A>G	p.K125K	5	SNP Finder	14	10	Yes-0.17	Yes
	c.598G>C	p.Q199H	7	BLAST	4	2	No-T81799	N.S.
				SNP Finder	1	1	No-0.01	N.S.
	c.669A>C	p.X223Y	7	BLAST	2	2	No-AI08445	N.S.
GSTA2	c.322C>T	p.L108F	5	BLAST	2	2	No-AI357144	N.S.
	c.335C>G	p.T112S	5	BLAST	9	3	Yes-AI357144	Yes
	c.384A>C	p.T128T	5	SNP Finder	2	2	No-0.05	N.S.
	c.389A>T	p.N130I	5	BLAST	4	3	No-AI357144	N.S.
	c.517A>G	p.S173G	6	BLAST	3	2	No-H71630	N.S.
	c.588G>T	p.K196N	7	BLAST	2	2	Yes-AI023858	No
				SNP Finder	2	2	Yes-0.96	No
	c.629A>C	p.E210A	7	BLAST	8	2	Yes-AI023858	Yes
	c.629C>A	p.A210E	7	SNP Finder	8	6	Yes-1.0	Yes

[a] The accession number of the I.M.A.G.E. Consortium (LLNL) cDNA clones sequenced to confirm the presence of potential polymorphisms identified using the BLAST alignment tool is provided.

[b] Polymorphisms detected using the SNP Finder program are allocated a likelihood score. As the number approaches 1.0, there is a greater likelihood that this polymorphism will be real. Sequences were confirmed by viewing the electopherogram data made available at the site.

[c] *GSTA1* was not represented in the UniGene clusters, hence was not able to be analysed using the SNP Finder program. Polymorphisms discovered in *GSTA1* were identified through the search performed for *GSTA2*, which downloaded several different contigs. Contig 4 corresponded to *GSTA2*, and Contig 5 to *GSTA1*.

N.S.: Not Studied

Table 3.2 – GST Alpha polymorphisms detected in the SNP databases.

Gene	Nucleotide	Residue	Exon	Database	Able to confirm[a]	Confirmed in population studies
GSTA1	c.115G>T	p.E39X[b]	3	HGBASE	Unable	No
	c.159G>T	p.Q53H[b]	4	refseq, HGBASE	Unable	No
	c.331G>T	p.V111L	5	dbSNP, refseq, GeneSNPs	Unable	No
	c.335G>C	p.C112S	5	dbSNP, refseq, GeneSNPs	Unable	N.S.
	c.382A>C	p.I128L	5	dbSNP, refseq, GeneSNPs	Unable	No
GSTA2	c.183T>C	p.D61D	4	EGPSNPs	Unable	N.S.
	c.269C>T	p.A90V	4	GeneSNPs	Unable	No
	c.306G>A	p.L102L	5	refseq, GeneSNPs	Unable	N.S.
	c.328C>T	p.P110S	5	EGPSNPs	Unable	Yes
	c.335G>C	p.S112T[b]	5	refseq (0.99)[c]	Yes	Yes
	c.446T>C	p.V149A	6	refseq, EGPSNPs, GeneSNPs	Unable	No
	c.588G>T	p.K196N[b]	7	dbSNP, refseq, GeneSNPs (0.96)[c], HGBASE	Yes	No
	c.629A>C	p.E210A[b]	7	dbSNP, refseq (0.99)[c], EGPSNPs, GeneSNPs (1.0)[c], HGBASE	Yes	Yes
GSTA3	c.211C>A	p.L71I	4	refseq, GeneSNPs	Unable	Yes
GSTA4	c.351G>A	p.Q117Q	5	dbSNP, refseq	Unable	N.S.

[a] Some sequences were unable to be confirmed directly through the databases, therefore small-scale population screens were used to confirm these polymorphisms.

[b] These polymorphisms were also detected by the BLAST alignment tool and the SNP Finder program.

[c] If the SNP was submitted by the GAI to different databases, links to the SNP Finder program were provided, allowing access to electropherograms.

Table 3.3 – Oligonucleotides used for *GSTA1*, *GSTA2* and *GSTA3* exon amplification.

Primer Name	Exon	Primer sequence 5' to 3'	Annealing Conditions	Product Length
GSTA1				
A1Ex3F	3	GAATGAACTAACAAGAACGT	50°C, 20 sec	124 bp
A1Ex3R	3	TCATCAGAGGAACTTAGAGA		
A1Ex4A	4	CACAACCATTTGTTCATCCCA	58°C, 18 sec	312 bp
A1Ex4B	4	ATCCTGCTGCTGGTCATGATG		
A1Ex4F2[a]	4	TGGATATTTGATG**A**T**G**CA	60°C, 18 sec	139 bp
A1Ex4R2	4	TACCGTACAGGGCTC		
A1Ex5F	5	GTTGTTTCTTGTCTTTCAGG	50°C, 18 sec	174 bp
A1Ex5R	5	CATCTTCACTTACTTTTTCA		
A1Ex5F2	5	CCCGTATGTCCACCTG	49°C, 18 sec	68 bp
A1Ex5HphI[b]	5	TAGCGATTTTTTATTTTC**A**C		
A1Ex5StuI[c]	5	AAATGATCCTCCTTC**A**G**G**CC	51°C, 18 sec	93 bp
A1Ex5DdeI[d]	5	CAGGGAAGTAGCGATTT**C**TTA		
GSTA2				
A2Ex4F	4	TTTCAAGTGAACTCTTACAGG	61°C, 18 sec	291 bp
A2Ex4MwoI[e]	4	AAGAACAGAAAATATA**G**CGTACAGG		
A2Ex5F	5	GTTGTTTTTTGTCTTTCAGG	55°C, 18 sec	174 bp
A2Ex5R	5	CAGCTTCACTTACTTTTTCA		
A2Ex5MwoI[f]		ATTTGGGTGAAATGAT**G**CTT	53°C, 16 sec	126 bp
A2Ex6F	6	TCTCCAGGTCTTAAAGAGCCAC	56°C, 18 sec	115 bp
A2Ex6R	6	AAGGCTAGAGTCAAGCTCTTCC		
A2Ex7F	7	TGTGCTTTGTGGATTACAGG	58°C, 20 sec	302 bp
A2Ex7R	7	CTAAGTGGGTGAATAGGAGT		
A2E7SphI[g]	7	AAGGAAGCCTCCCATG**C**ATG	55°C, 18 sec	221 bp
GSTA3				
A3Ex4F	4	CATTTTATAACCTCAGTCATTTCAACCATC	57°C, 18 sec	327 bp
A3Ex4R		CCTGGTCATGATGCCCTGTCATGGTCT		

[a] A modification of ttc → atg (underlined) was introduced to create an *Nsi*I site in amplified DNA containing the p.Q53H variation. This primer was used in combination with A1Ex4R2 on PCR products produced by A1Ex4A and A1Ex4B.

[b] A modification of ctc → cac (underlined) was introduced to create an *Hph*I site in amplified DNA containing the c.365G>A silent variation. This primer was used in combination with A1Ex5F2 on PCR products produced by A1Ex5F and A1Ex5R.

[c] A modification of tgc → agg (underlined) was introduced to create a *Stu*I site in amplified DNA containing the p.V111L variation. This primer was used in combination with A1Ex5DdeI.

[d] A modification of t → c (underlined) was introduced to create a *Dde*I site in amplified DNA containing the p.C112S variation. This primer was used in combination with A1Ex5StuI.

[e] A modification of c → g (underlined) was introduced to create an *Mwo*I site in amplified DNA containing the p.A90V variation.

[f] A modification of c → g (underlined) was introduced to create an *Mwo*I site in amplified DNA containing the p.P110S variation. This primer was used in conjunction with A2Ex5R.

[g] A modification of g → c (underlined) was introduced to create an *Sph*I site in amplified DNA containing the p.E210A variation. This primer was used in combination with A2Ex7R.

Table 3.4 – Restriction Endonuclease Sites used to detect polymorphisms in the *GSTA1*, *GSTA2* and *GSTA3* genes.

Variation	Exon	Endonuclease Site	Fragment Size (bp)		
			Wild-type	Heterozygous	Variant
GSTA1					
p.E39X	3	*Pst*I	68, 55	123, 68, 55	123
p.Q53H	4	*Nsi*I	139	139, 121, 18	121, 18
p.V111L	5	*Stu*I	92	18, 75, 92	18, 75
p.K125K	5	*Hph*I	68	68, 60, 8	60, 8
p.I128L	5	*Dde*I	31, 62	18, 61, 44, 62	18, 31, 44
GSTA2					
p.A90V	4	*Mwo*I	20, 271	20, 271, 291	291
p.P110S	5	*Mwo*I	20, 37, 69	20, 37, 57, 69	57, 69
p.T112S	5	*Mbo*II/*Hinc*II	114, 60	114, 90, 60, 24	90, 60, 24
p.V194A	6	*Cac*8I	114	40, 74, 114	40, 74
p.K196N	7	*Apo*I	302	302, 244, 58	244, 58
p.E210A	7	*Sph*I	221	221, 201, 20	201, 20
GSTA3					
p.L71I	4	*Bsp*1286I	60, 99, 168	60, 99, 168, 228	99, 228

3.2.3 Recombinant Protein Work

3.2.3.1 GSTA2-2

3.2.3.1.1 Site-Directed Mutagenesis

The GSTA2-2 protein encoded by the I.M.A.G.E. Consortium (LLNL) cDNA Clone T96198 was expressed in *E. coli* as an ubiquitin fusion in the plasmid pRB269 [Baker *et al.*, 1994]. The T96198 cDNA was amplified by PCR using the forward primer 5'-GACCGCGGTGGTATGGCAGAGAAGCCCAAGC-3' and the reverse primer 5'-ACGTCGACTCATTAAAACCTGAAAATCTTCC-3' to create a *Sac*II restriction site upstream of the GSTA2-2 initiation codon, and a *Sal*I site downstream of the GSTA2-2 stop codon. The 692 bp amplified product was digested with *Sac*II and *Sal*I and the resulting fragment ligated into the *Sac*II and *Sal*I sites of pRB269. This construct was confirmed by sequencing as being the GSTA2D protein (Table 3.5) and was created by Dr Dan Liu. Using the QuikChange Site-Directed Mutagenesis Kit (§2.6.3), the GSTA2B protein (containing the p.A210 variation) was created from GSTA2D using the primer pair A27KN1 and A27KN2 (Table 3.6).

The GSTA2A protein was constructed (using cloning and ligation techniques outlined in §2.5.4.1 and §2.6.1) by replacing the 210 bp *Dra*III-*Sph*I fragment in the GSTA2D protein with the 215 bp *Dra*III-*Nsp*I fragment from the I.M.A.G.E. Consortium (LLNL) cDNA Clone AI357144. The GSTA2C and GSTA2E proteins (containing the p.T112 and p.S110 variations respectively) were created from the GSTA2A protein using the QuikChange Site-Directed Mutagenesis Kit (§2.6.3) and the primer pairs A25ST1 and A25ST2, and GSTA2P110SA and GSTA2P110SB respectively (Table 3.6). All mutations and constructs were confirmed by DNA cycle sequencing (§2.5.5.1).

3.2.3.1.2 Protein Expression and Purification

Plasmids encoding the GSTA2A, GSTA2B and GSTA2C variants were transfected into *E. coli* strains XL1-Blue or TG1 (§2.6.2) containing the plasmid pRB173 [Baker *et al.*,

Table 3.5 – Nucleotide and amino acid variations in *GSTA2*.

Haplotype	**Nucleotide**			
	c.328	c.335	c.588	c.629
GSTA2A	c	g	g	a
GSTA2B	c	g	g	c
GSTA2C	c	c	g	a
GSTA2D	c	g	t	c
GSTA2E	t	g	g	a
	Residue			
	p.110	p.112	p.196	p.210
GSTA2A	Pro	Ser	Lys	Glu
GSTA2B	Pro	Ser	Lys	Ala
GSTA2C	Pro	Thr	Lys	Glu
GSTA2D	Pro	Ser	Asn	Ala
GSTA2E	Ser	Ser	Lys	Glu

Table 3.6 – Mutagenesis primers used to create GSTA2-2 recombinant protein variants.

GSTA2 Haplotype	Primer Name	Mutagenesis Primer Sequences 5' to 3'
GSTA2B	A27KN1	CCCACAGTGAAGAAGTTTCTACAGCCTGGC
	A27KN2	GCCAGGCTGTAGAAACTTCTTCACTGTGGG
GSTA2C	A25ST1	CTTCTGCCCTTTACTCAACCTGAGGAAC
	A25ST2	GTTCCTCAGGTTGAGTAAAGGGCAGAAG
GSTA2E	GSTA2P110SA	GGTGAAATGATCCTTCTTCTGTCCTTTAGTCAACCTGAGGAAC
	GSTA2P110SB	GTTCCTCAGGTTGACTAAAGGACAGAAGAAGGATCATTTCACC

1994], a source of the ubiquitin-specific protease Ubp2 [Baker *et al.*, 1992]. This ubiquitin specific protease co-translationally removes the N-terminal ubiquitin moiety from the fusion proteins. 4 mL of an overnight culture was diluted 1:100 into LB supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. The 400 mL cultures were incubated at 37°C with vigorous shaking to an OD_{600} of 0.6 and protein expression was induced with 0.5 mM IPTG for 3 hours. The cells were harvested by centrifugation at 5 K for 10 minutes at 4°C (GSA-1500 rotor), resuspended in 15 mL of 20 mM NaPi, pH 7.0 and lysed by sonication (5 x 15 second bursts). Cellular debris was removed by centrifugation at 12 K for 20 minutes at 4°C (SS-34 rotor).

All proteins were purified at 4°C by affinity chromatography. Cleared lysates were diluted with a further 10 mL of 20 mM NaPi, pH7.0 and incubated with GSH agarose at 4°C for 1 hour on a rotary mixer. The agarose was collected by centrifugation at 1000 rpm for 5 minutes and washed twice with 40 mL of 20 mM NaPi, pH 7.0. The agarose was washed with 200 mL of 20 mM NaPi, pH 7.0 on a scintered funnel, then transferred to a column and washed with a further 50 mL of 50 mM Tris-HCl, pH 7.2. The pure protein was eluted from the GSH-agarose with 50 mM GSH in 200 mM Tris-HCl, pH 9.6. Fractions containing recombinant GSTA2 protein were identified by 12.5% SDS-PAGE and stored in 10% glycerol at -20°C. Protein concentrations were determined as described in §2.5.2.3.

3.2.3.2 GSTA3-3

3.2.3.2.1 Amplification and Isolation of the *GSTA3* cDNA

cDNA was extracted from existing lambda phage lysate prepared from an LE392:λgt11 human testis cDNA library as described in §2.5.1.1. The *GSTA3* cDNA was amplified from this cDNA using the primer pair GSTA3EcoRIF2 5'-AATAATGAATTCATGGCAGGGAAGCCCAAGCTT and GSTA3NCBamHIR 5'-AATAATGGATCCTTCTTAGCCTCCATGGCTGCT [Johansson & Mannervik, 2001a], which have an *Eco*RI and *Bam*HI restriction site respectively incorporated. A 20 µL reaction volume contained 1x cloned *Pfu* 10x reaction buffer, 0.2 mM dNTPs,

1.5 mM $MgCl_2$, 6 pmol of the forward and reverse primers and 0.5 unit of *Pfu* Turbo DNA polymerase. A negative control sample containing all components of the reaction except the template was included in each PCR to ensure the absence of contamination. PCR conditions involved an initial denaturation step at 95°C for 5 minutes followed by 35 cycles of a 95°C denaturation step (2 min), 54°C annealing step (1 min), and 72°C extension step (1 min), with a final 7 minute extension at 72°C [Johansson & Mannervik, 2001a]. Agarose gel electrophoresis was used to confirm the absence of spurious bands and contamination and the presence of PCR product (5 µL).

3.2.3.2.2 Subcloning

The 667 bp and 719 bp GSTA3 PCR products were digested with the restriction enzymes *Eco*RI and *Bam*HI and cloned into the *Bam*HI and *Eco*RI sites of the pKK263 vector as described in §2.5.4.1 and §2.6.1. The constructs were confirmed by automated sequencing as described in §2.5.5.2.

3.2.4 GSTA2-2 Enzyme Characterisation

3.2.4.1 Structural Analysis

The GSTA2A [p.P110;S112;E210], GSTA2B [p.P110;T112;E210], GSTA2C [p.P110;S112;A210] and GSTA2E [p.S110;S112;E210] variants were built into a homology model of GSTA2-2 based on the co-ordinates of GSTA1-1 (PDB code 1GSE) as described in §2.6.5.

3.2.4.2 Enzymatic Analysis

Glutathione *S*-transferase activity towards a range of typical Alpha class substrates was measured as described in §2.6.4. Three separate preparations of the GSTA2A, GSTA2B and GSTA2C isoforms were prepared. Two separate preparations of the

GSTA2A and GSTA2E isoforms were prepared at a later stage. No changes in the yield or expression levels of these variants were observed.

3.2.4.3 Steady State Kinetic Analysis

Steady state kinetics for the recombinant proteins GSTA2A and GSTA2E against the substrate CDNB were determined using a 5x5 experiment. Kinetic parameters for the CDNB reaction were determined by measuring the kinetics for five concentrations of CDNB (1000, 800, 600, 400 and 200 μM) against five concentrations of GSH (1000, 500, 200, 100 and 50 μM GSH). Similarly, kinetic parameters for the *p*-nitrophenylacetate (*p*NPA) assay were determined by measuring the kinetics for five concentrations of *p*NPA (1000, 750, 400, 200 and 100 μM) against five concentrations of GSH (1000, 500, 200, 100 and 50 μM GSH). The steady state kinetic parameters were determined by fitting the Michaelis Menten equation to the data points using the MacCurve Fit program. Hyperbolae were fitted to bring R^2 closest to 1 ($\leq$1), and SSE closest to 0.0001 ($\geq$0.0001).

3.3 Results

3.3.1 Detection of Alpha Class GST Polymorphisms

3.3.1.1 Detection using Database Analysis Programs

The BLAST alignment tool and the SNP Finder program were used to detect polymorphisms in the Alpha class GST cDNA sequences located in the human EST database and UniGene clusters respectively. The EST database contained 136 *GSTA1*, 46 *GSTA2*, 20 *GSTA3* and 95 *GSTA4* cDNAs. The UniGene clusters contained 132 *GSTA2* (of which 99 were subsequently verified as *GSTA1* and 33 *GSTA2*), eight *GSTA3* and one *GSTA4* cDNA and mRNA sequences. From these sequence databases, eleven variants were detected in the *GSTA1* and *GSTA2* genes (Table 3.1). Of these, only six variants were confirmed upon resequencing a representative I.M.A.G.E. Consortium (LLNL) cDNA clone or viewing existing electropherographic data.

3.3.1.2 Detection using SNP Databases

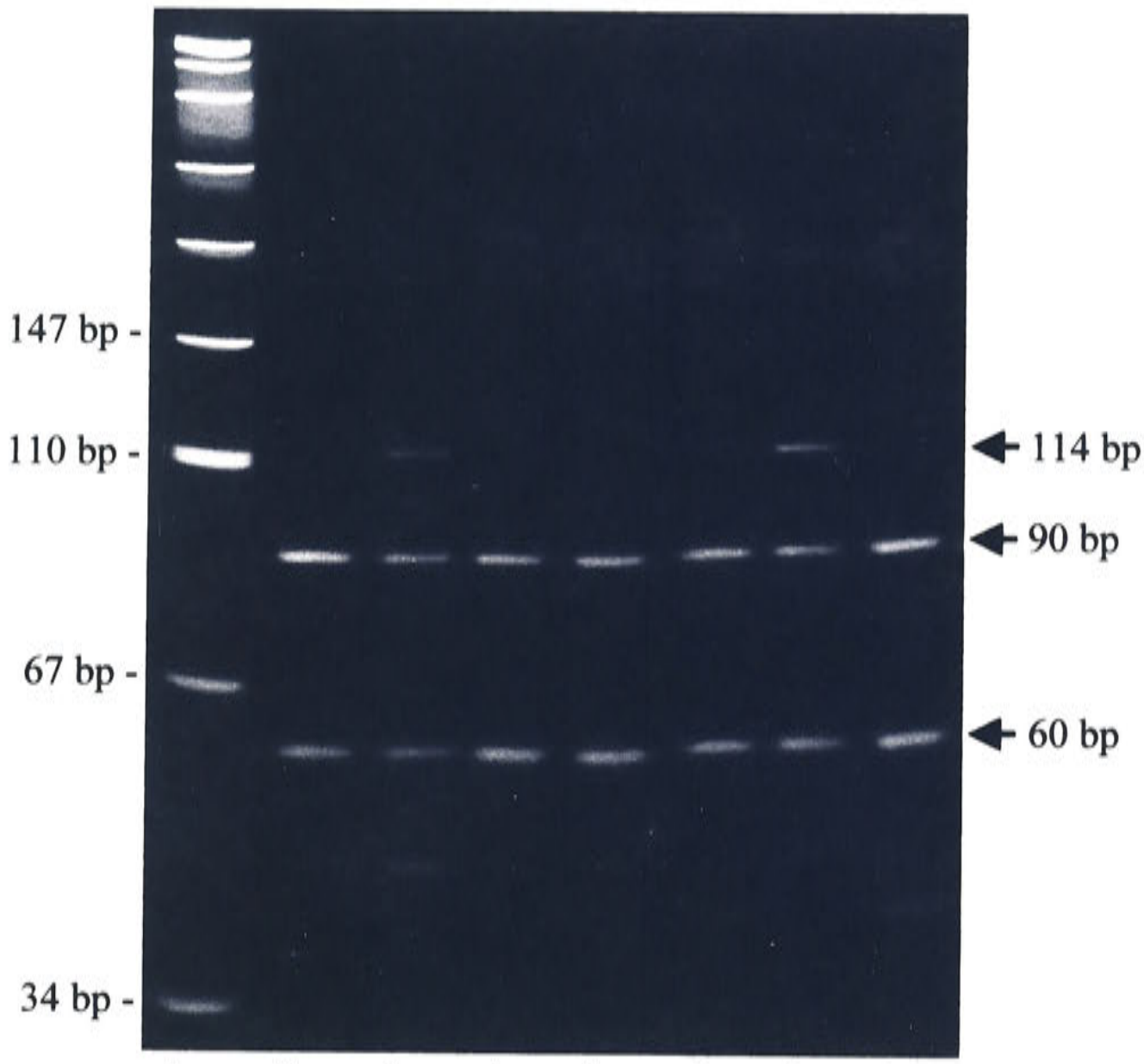
A number of SNP databases were also analysed for GST Alpha class submissions. Multiple variants were listed in these databases, however when limited to those that altered amino acid residues in the coding region, seven novel variants were found: the p.V111L, p.C112S and p.I128L substitutions in *GSTA1*; the p.A90V, p.P110S and p.V149A substitutions in *GSTA2*; and the p.L71I substitution in the *GSTA3* gene (Table 3.2). Although some *GSTA2* variants could be confirmed through analysis of electropherograms and available frequency data, small-scale population analysis was undertaken for confirmation on all polymorphisms identified. Four variants that had been detected using the BLAST alignment tool and SNP Finder programs were also detected in these databases.

3.3.1.3 Determination and Distribution of Allele Frequencies

Amplification of the variant exons followed by digestion with the appropriate restriction enzyme allowed simple determination of the nucleotide present at each of the 13 potential variant sites. For the six variants detected using the BLAST alignment tool and SNP Finder programs, 200 Australian Europeans, 87 Creole Africans and 97 Southern Chinese samples were initially screened. Only three polymorphisms were identified and confirmed in these populations: the silent p.K125K (c.365G>A) *GSTA1* variant, and the p.T112S (Figure 3.1) and p.E210A (Figure 3.2) *GSTA2* variants (Table 3.7). 74 Bantu African samples were subsequently screened for the two *GSTA2* variants. In *GSTA1*, the silent polymorphism was commonly found in all populations studied, but the c.G365 allele was rarer in the Creole African and Southern Chinese populations. Chi-square analysis provided support for a difference in genotypic distribution between the three races ($\chi^2_4 = 54.74$, $p<0.05$). In *GSTA2*, the p.T112 variant of the p.T112S polymorphism was only ever detected in the heterozygous state. This variant was found to be extremely rare in all but the Bantu African population, in which the frequency for this allele increased significantly from the 0%-0.5% observed in the Australian European, Creole African and Southern Chinese groups to 15%. Chi-square analysis provided support for a difference in genotypic distribution between these populations ($\chi^2_6 = 102.48$, $p<0.05$), specifically between the Bantu Africans and the remaining three populations (Australian European/Bantu African: $\chi^2_2 = 56.62$, $p<0.05$; Creole African/Bantu African: $\chi^2_2 = 27.10$, $p<0.05$; Southern Chinese/Bantu African: $\chi^2_2 = 33.58$, $p<0.05$). In contrast, the p.A210 variant of the p.E210A polymorphism occurred quite frequently, ranging from an extremely high allele frequency of 42% in the Bantu African population to a mere 8% in the Australian European population. Chi-square analysis provided support for a difference in genotypic distribution between the four populations ($\chi^2_6 = 124.54$, $p<0.05$). Further analysis demonstrated no difference in genotypic distribution between the Creole

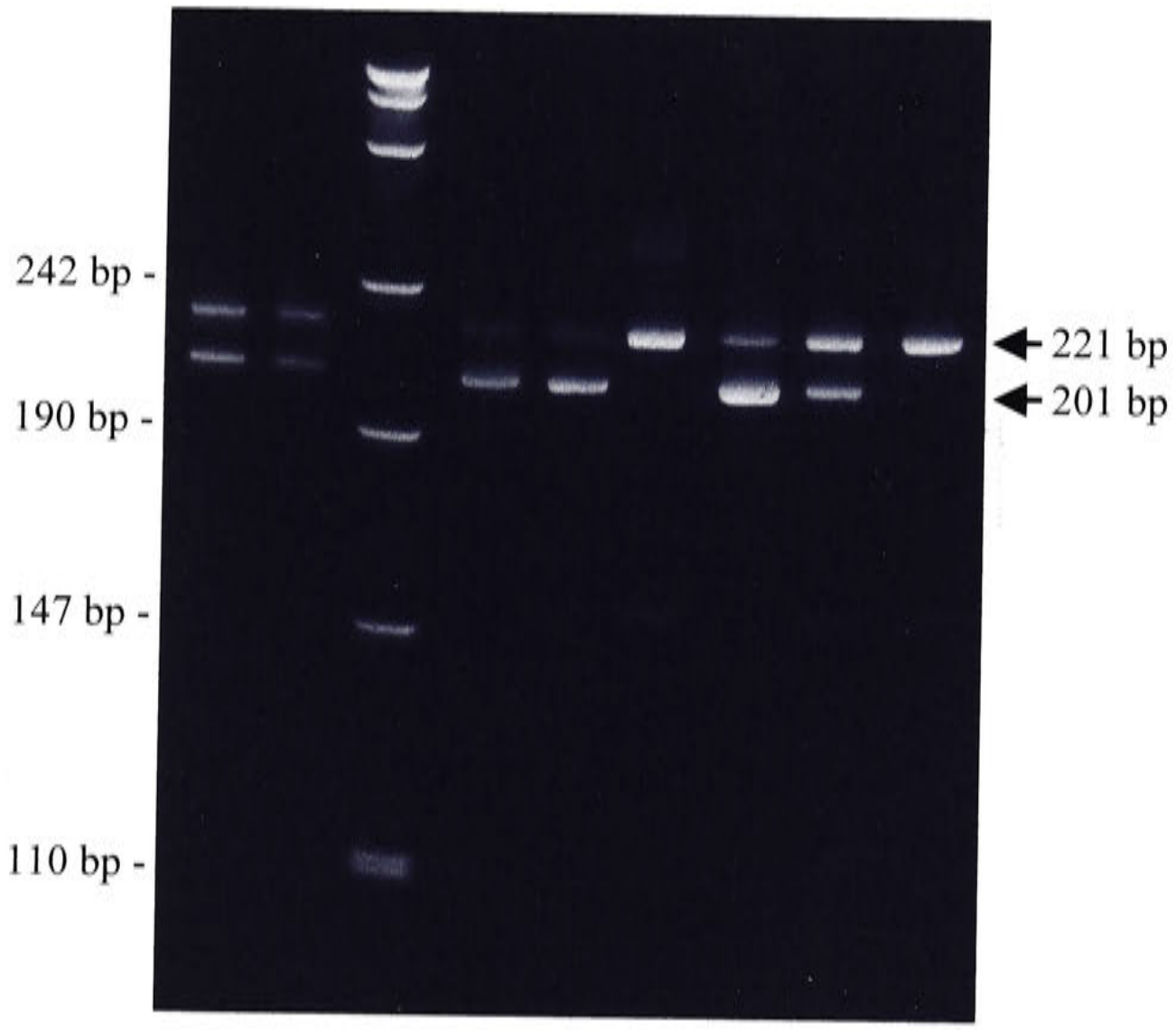
5.47, $p>0.05$). ***Of note, this polymorphism was not in accordance with the Hardy-Weinberg in the Bantu African population. It is possible that this locus is in linkage disequilibrium with other loci."***

population screening: the *GSTA2* p.T112S (Figure 3.3) and the *GSTM3* p.E27A (Figure 3.4) polymorphisms. Screening additional samples from each of the three populations,



Fragment (bp)	C/C	C/G	G/G
114	—	—	
90		—	—
60	—	—	—
24		—	—

Figure 3.1 - PCR/RFLP analysis of the *GSTA2* p.T112S variant. A 174 bp PCR product spanning exon 5 of the *GSTA2* gene was digested with *Hinc*II and *Mbo*II for nucleotide 335 determination. Lane 1 - pUC/*Hpa*II marker; Lanes 2, 4-6, 8 - G/G homozygotes; Lanes 3, 7 - C/G heterozygotes.



Fragment (bp)	A/A	A/C	C/C
221	—	—	
201		—	—
20		—	—

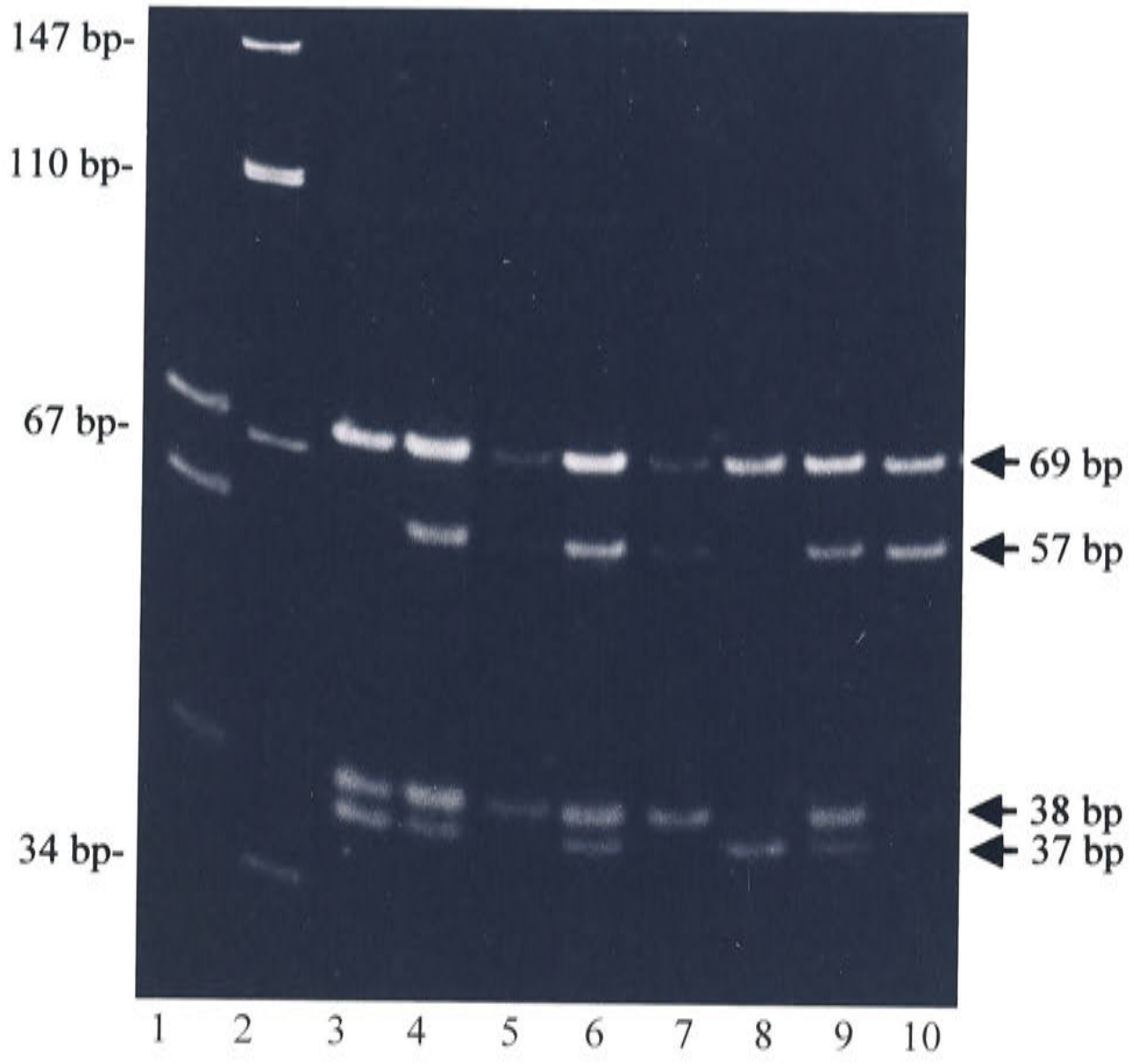
Figure 3.2 - PCR/RFLP analysis of the *GSTA2* p.E210A variant. A 221 bp PCR product spanning exon 7 of the *GSTA2* gene was digested with *Sph*I for nucleotide 629 determination. Lanes 1, 2, 8 - A/C heterozygotes; Lanes 4, 5, 7 - C/C homozygotes; Lanes 6, 9 - A/A heterozygotes; Lane 3 - pUC/*Hpa*II marker.

Table 3.7 – GST Alpha allele and haplotype frequencies in three ethnic groups.

Gene	Polymorphism	Population	n	Genotype			Allele Frequency	
GSTA1	p.K125K	Australian	200	G/G=40	G/A=86	A/A=74	c.G365=0.42	c.A365=0.58
		Creole African	87	G/G=4	G/A=38	A/A=45	c.G365=0.26	c.A365=0.74
		Chinese	96	G/G=2	G/A=19	A/A=75	c.G365=0.12	c.A365=0.88
GSTA2	p.P110S	Australian	192	C/C=173	C/T=19	T/T=0	p.P110=0.95	p.S110=0.05
		Bantu African	65	C/C=63	C/T=2	T/T=0	p.P110=0.98	p.S110=0.21
		Creole African	80	C/C=61	C/T=17	T/T=2	p.P110=0.87	p.S110=0.13
		Chinese	100	C/C=78	C/T=20	T/T=2	p.P110=0.88	p.S110=0.12
	p.T112S	Australian	200	C/C=0	C/G=2	G/G=198	p.T112=0.005	p.S112=0.995
		Bantu African	73	C/C=0	C/G=22	G/G=51	p.T112=0.15	p.S112=0.85
		Creole African	87	C/C=0	C/G=1	G/G=86	p.T112=0.006	p.S112=0.994
		Chinese	97	C/C=0	C/G=0	G/G=97	p.T112=0	p.S112=1
	p.E210A	Australian	200	A/A=169	A/C=30	C/C=1	p.E210=0.92	p.A210=0.08
		Bantu African	71	A/A=12	A/C=59	C/C=0	p.E210=0.58	p.A210=0.42
		Creole African	87	A/A=43	A/C=39	C/C=5	p.E210=0.72	p.A210=0.28
		Chinese	97	A/A=61	A/C=35	C/C=1	p.E210=0.81	p.A210=0.19
GSTA3	p.L71I	Australian	100	C/C=0	C/A=0	A/A=100	p.L71=0	p.I71=1
		Bantu African	73	C/C=1	C/A=20	A/A=52	p.L71=0.15	p.I71=0.85
		Creole African	85	C/C=1	C/A=6	A/A=78	p.L71=0.05	p.I71=0.95
		Chinese	97	C/C=0	C/A=0	A/A=97	p.L71=0	p.I71=1

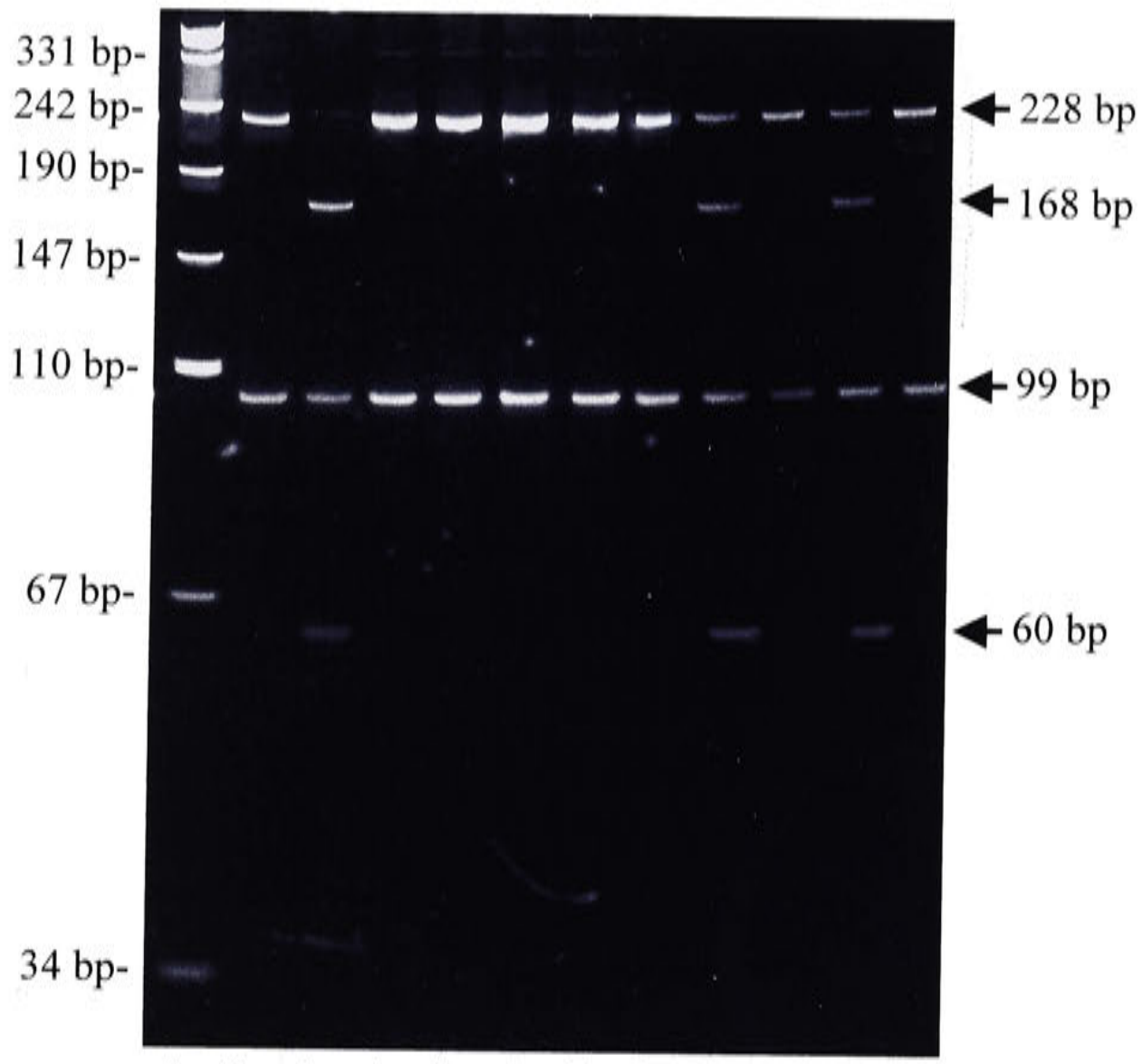
Gene	Population	n	Haplotype			
			GSTA2A [p.P110;S112;E210]	GSTA2B [p.P110;S112;A210]	GSTA2C [p.P110;T112;E210]	GSTA2E [p.S110;S112;E210]
GSTA2	Australian	192	0.86	0.08	0.0052	0.049
	Bantu African	64	0.40	0.43	0.16	0.02
	Creole African	80	0.60	0.27	-	0.13
	Chinese	89	0.70	0.19	-	0.11

Note: All polymorphisms are in agreement with the Hardy-Weinberg equilibrium, with the exception of the p.E210A polymorphism in the Bantu African population.



Fragment (bp)	C/C	C/T	T/T
69	—	—	—
57		—	—
37	—	—	
20	—	—	

Figure 3.3 - PCR/RFLP analysis of the *GSTA2* p.P110S variant. A 126 bp PCR product spanning exon 5 of the *GSTA2* gene was digested with *Mwo*I for nucleotide 328 determination. Lanes 1, 7 - T/T homozygotes; Lanes 3, 8 - C/C homozygotes; Lanes 4-6, 9, 10 - C/T heterozygotes; Lane 2 - pUC/*Hpa*II marker. The 38 bp fragment is not associated with the *Mwo*I digestion, but is primer dimer associated with the PCR product.



Fragment (bp)	C/C	C/A	A/A
228		—	—
168	—	—	
99	—	—	—
60	—	—	

Figure 3.4 - PCR/RFLP analysis of the *GSTA3* p.L71I variant. A 327 bp PCR product spanning exon 4 of the *GSTA3* gene was digested with *Bsp*1286I for nucleotide 211 determination. Lanes 2, 4-8, 10, 12 - A/A homozygotes; Lane 3 - C/C homozygote; Lanes 9, 11 - C/A heterozygotes; Lane 1 - pUC/*Hpa*II marker.

plus 74 samples from the Bantu African population, to determine allele frequencies for these polymorphisms revealed that the *GSTA2* p.S110 variant was relatively uncommon, especially in the Australian European and Bantu African populations. The p.S110 allele frequency was higher in the Creole African and Southern Chinese populations at approximately 11%-12%, due to the identification of four homozygous individuals. Chi-square analysis provided support for a genotypic difference between the four populations ($\chi^2_6 = 22.73$, $p<0.05$). Further analysis demonstrated that there was no genotypic difference between the Australian European and Bantu population ($\chi^2_2 = 3.01$, $p<0.05$). The *GSTA3* p.L71 variant was found to be even more uncommon, appearing only in the two African populations where only one individual from each group was homozygous for the allele. Again, Chi-square analysis provided support for a genotypic difference between the four populations ($\chi^2_6 = 61.90$, $p<0.05$).

The occurrence of alternate alleles at different nucleotide positions in GSTA2-2 results in the generation of 16 possible haplotype combinations, although evidence for only five of these was obtained in this study. These five haplotypes have been defined in Table 3.5. Four of these were observed in the Australian European and Bantu African populations, and only three in the Creole African and Chinese populations, based on the assumption that the combination of p.S110 with the p.A210 residue is never observed (Table 3.7). This assumption was based on the observation that the four p.S110 homozygotes detected were also homozygous for the p.E210 variant, as were 47 of the 58 p.S110 heterozygotes. The remaining 11 p.S110 heterozygotes were heterozygous for the p.E210/A210 variant, hence it was therefore assumed that the p.S110 residue was associated with the p.E210 residue. The haplotype defined as GSTA2D (Table 3.5) was identified in the I.M.A.G.E. Consortium (LLNL) cDNA clone T96198, but was not observed in the three populations studied here.

3.3.2 Enzyme Characterisation

3.3.2.1 GSTA2-2

3.3.2.1.1 Structural Analysis

To evaluate the potential effects of the p.P110S, p.T112S and p.E210A variants of GSTA2-2, the substitutions were modelled into the 3-D structure of GSTA1-1 (Figure 3.5). The GSTA1-1 p.K125K polymorphism was not analysed further. Being a silent polymorphism, p.K125K is not expected to have any effects upon the function of the protein. Residue 110 is located in the H-site of the GSTA2-2 protein (Figure 3.6) at the C-terminus of the α4 helix (Figure 3.5), and although the p.P110S change is non-conservative – proline is a nonpolar residue and serine an uncharged polar residue, predictive modelling demonstrates that this substitution only introduces minor changes to the overall protein structure (Figure 3.7A). However, the substitution of p.S110 for p.P110 introduces one new H-bond between p.L106 and p.L108 (Figure 3.7B). Neither residue 112, located in the loop between the α4 and α5 helices, nor residue 210, which is located at the beginning of the C-terminal α9 helix, interacts with the active site. Predictive modelling indicates that these two residues do not alter the protein structure (Figures 3.8A and 3.9A). Although the p.T112S change is conservative, both serine and threonine are polar non-charged residues, the substitution of p.T112 for p.S112 introduces one new H-bond between p.T112 and p.L108 (Figure 3.8B). The p.E210A substitution is non-conservative, with alanine, a small polar residue being replaced by glutamic acid, a larger, acidic residue. Despite being a non-conservative substitution, only one H-bond, between p.D209 and p.E210, is lost upon the replacement of p.E210 with p.A210 (Figure 3.9B).

3.3.2.1.2 Enzymatic Analysis

Despite the structural modelling indicating that the p.T112 and p.A210 variants were unlikely to have significant functional effects, the activities of these variants were examined with typical Alpha class substrates (Table 3.8). As predicted, there were no

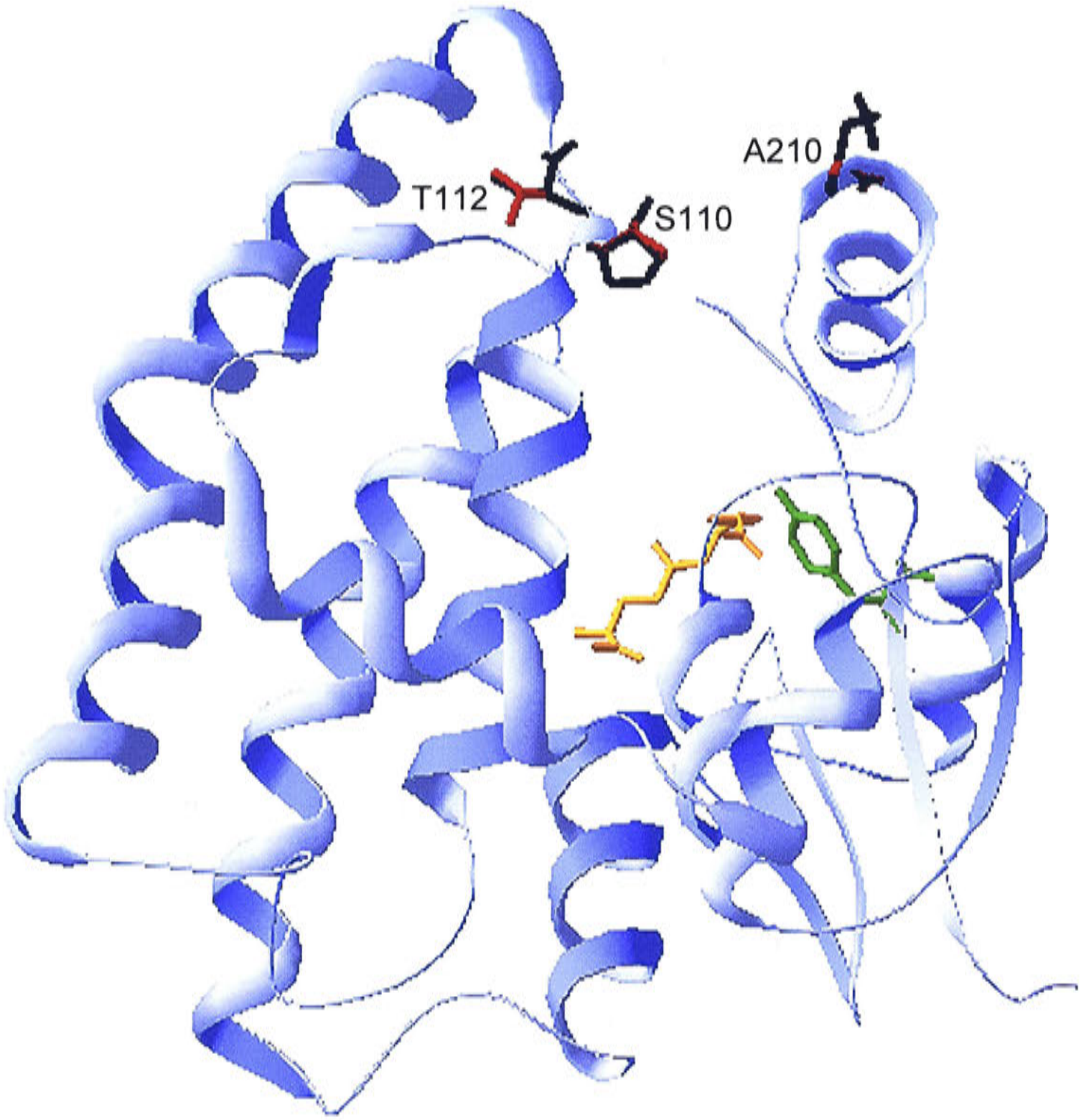
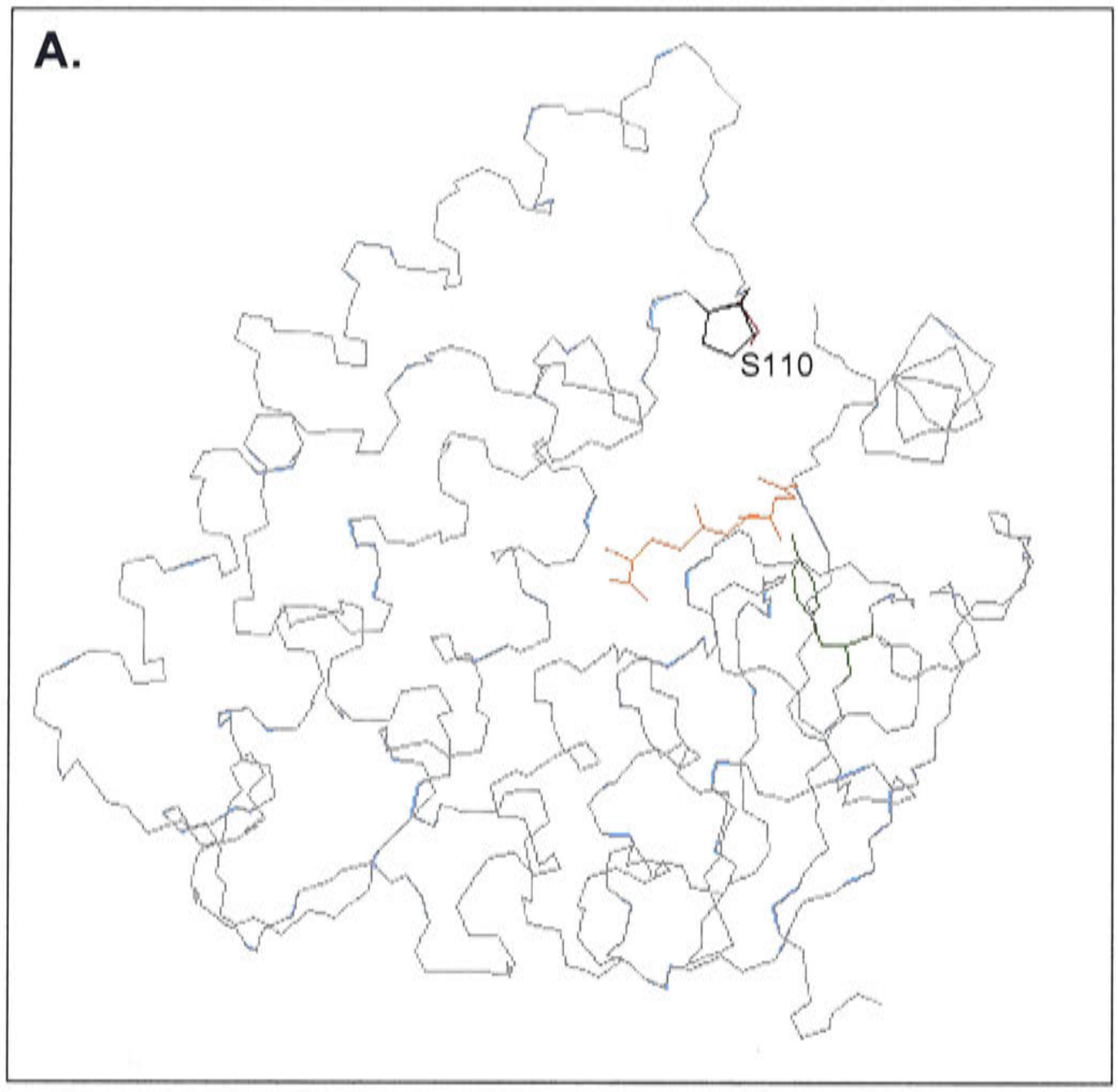


Figure 3.5 – Ribbon diagram of GSTA2 depicting the position of the polymorphic p.P110S, p.T112S and p.E210A substitutions. The most common amino acids, p.P110, p.S112 and p.E210, are shown in black. The least common residues, p.S110, p.T112 and p.A210 are coloured red. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.Y9 residue is coloured green.

Figure 3.6 – Protein sequence alignment of human GSTA1-1, GSTA2-2, GSTA3-3 and GSTA4-4. Identity with the corresponding residue in GSTA1-1 is indicated with a dot. Residues highlighted in blue are situated in the G-site, and those highlighted in red in the H-site, according to the crystal structure of GSTA1-1 [Sinning *et al.*, 1993]. Adapted from Figure 2 in Morel *et al.*, 2002.

Figure 3.7 – Carbon backbone of the GSTA2 monomer encoding p.P110 overlaid with the protein encoding p.S110. GSTA2 encoding the more common amino acid p.P110 (depicted in black) is coloured grey; GSTA2 encoding the less common amino acid p.S110 (depicted in red) is coloured blue. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.Y9 residue is coloured green. (A). Substitution of the labelled p.S110 residue for the p.P110 residue introduces minor changes to the overall structure of the GSTA2 protein. (B). Magnification of residue 110 and neighbouring residues showing the different H-bond patterns. H-bonds common to both variants are coloured dark blue; those introduced by the substitution of p.S110 for p.P110 are shown in pink.


A.
S110

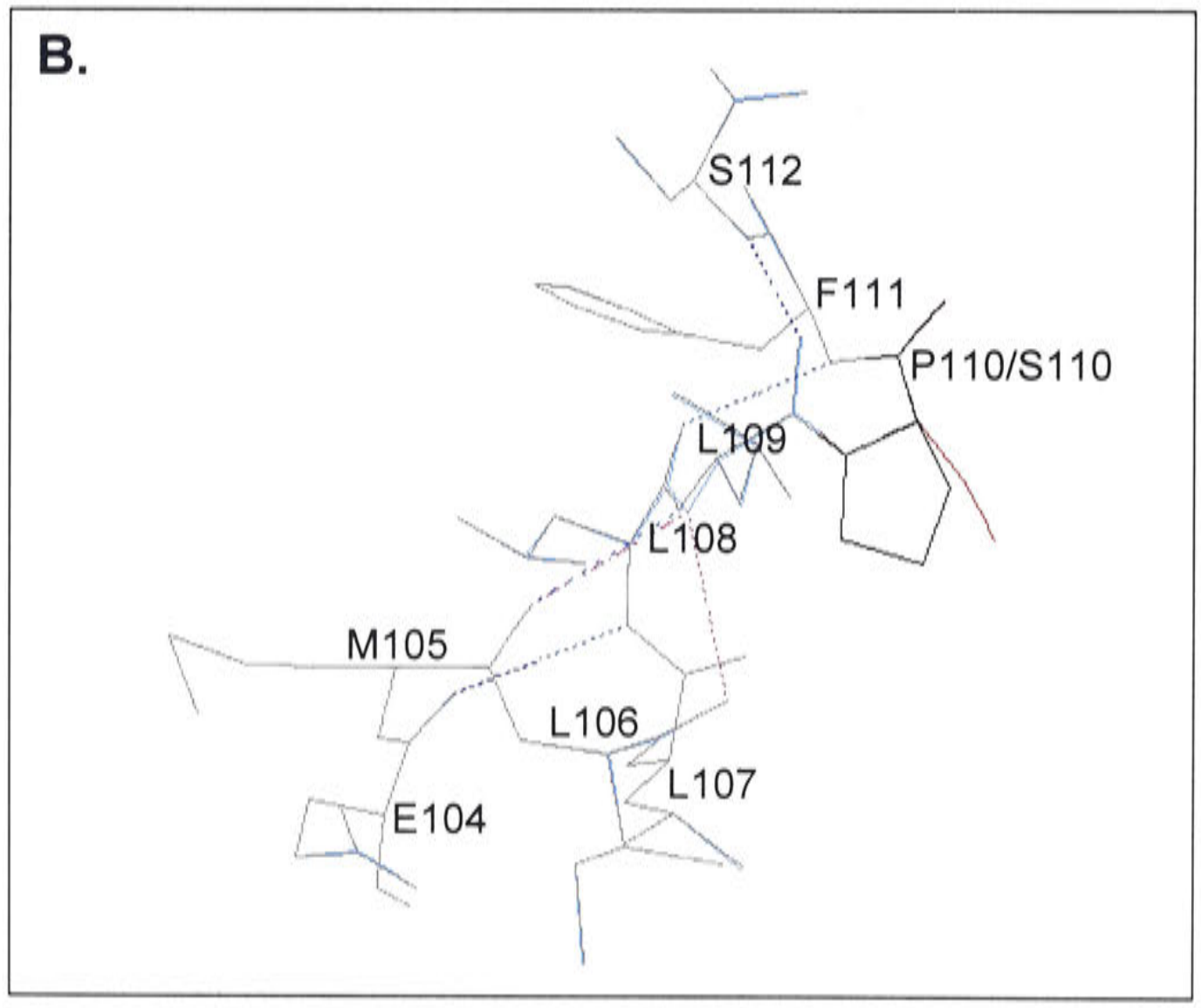
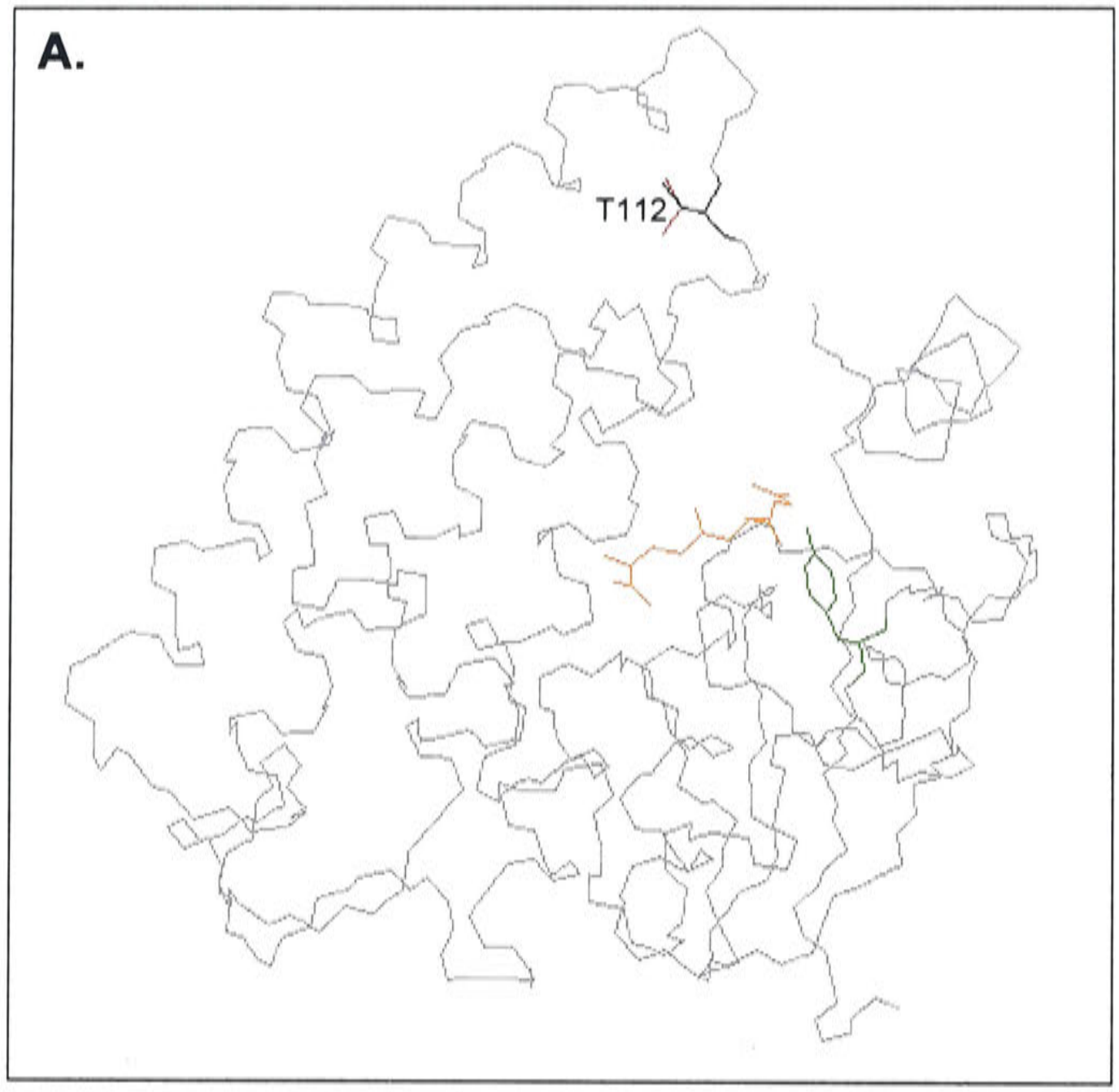

B.
S112
F111
P110/S110
L109
L108
M105
L106
L107
E104

Figure 3.8 – Carbon backbone of the GSTA2 monomer encoding p.S112 overlaid with the protein encoding p.T112. GSTA2 encoding the more common amino acid p.S112 (depicted in black) is coloured grey; GSTA2 encoding the less common amino acid p.T112 (depicted in red) is coloured blue. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.Y9 residue is coloured green. (A). Substitution of the p.S112 residue for the labelled p.T112 residue does not alter the overall structure of the GSTA2 protein. (B). Magnification of residue 112 and neighbouring residues showing the different H-bond patterns. H-bonds common to both variants are coloured dark blue; those introduced by the substitution of p.T112 for p.S112 are shown in pink.


A.
T112

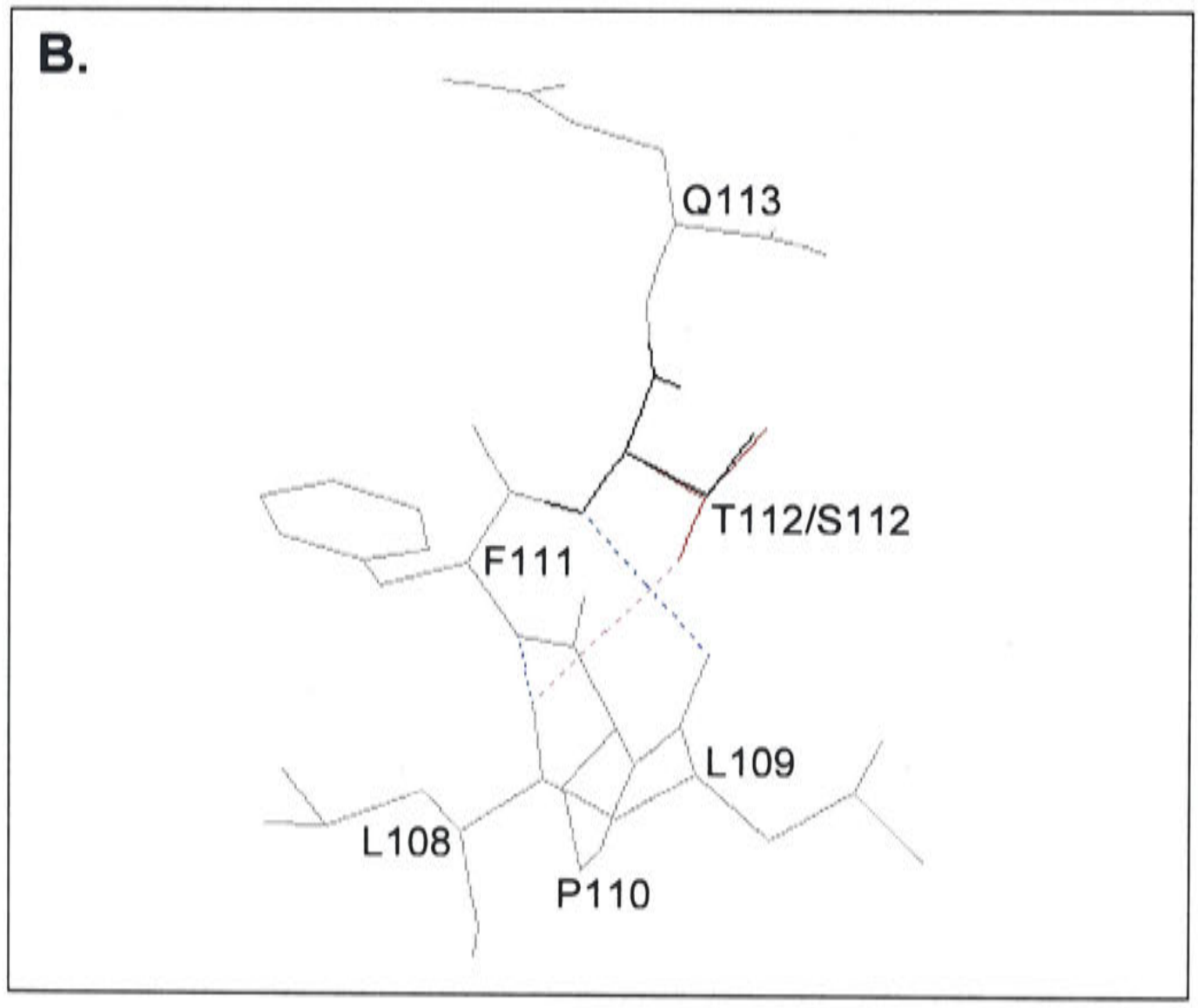
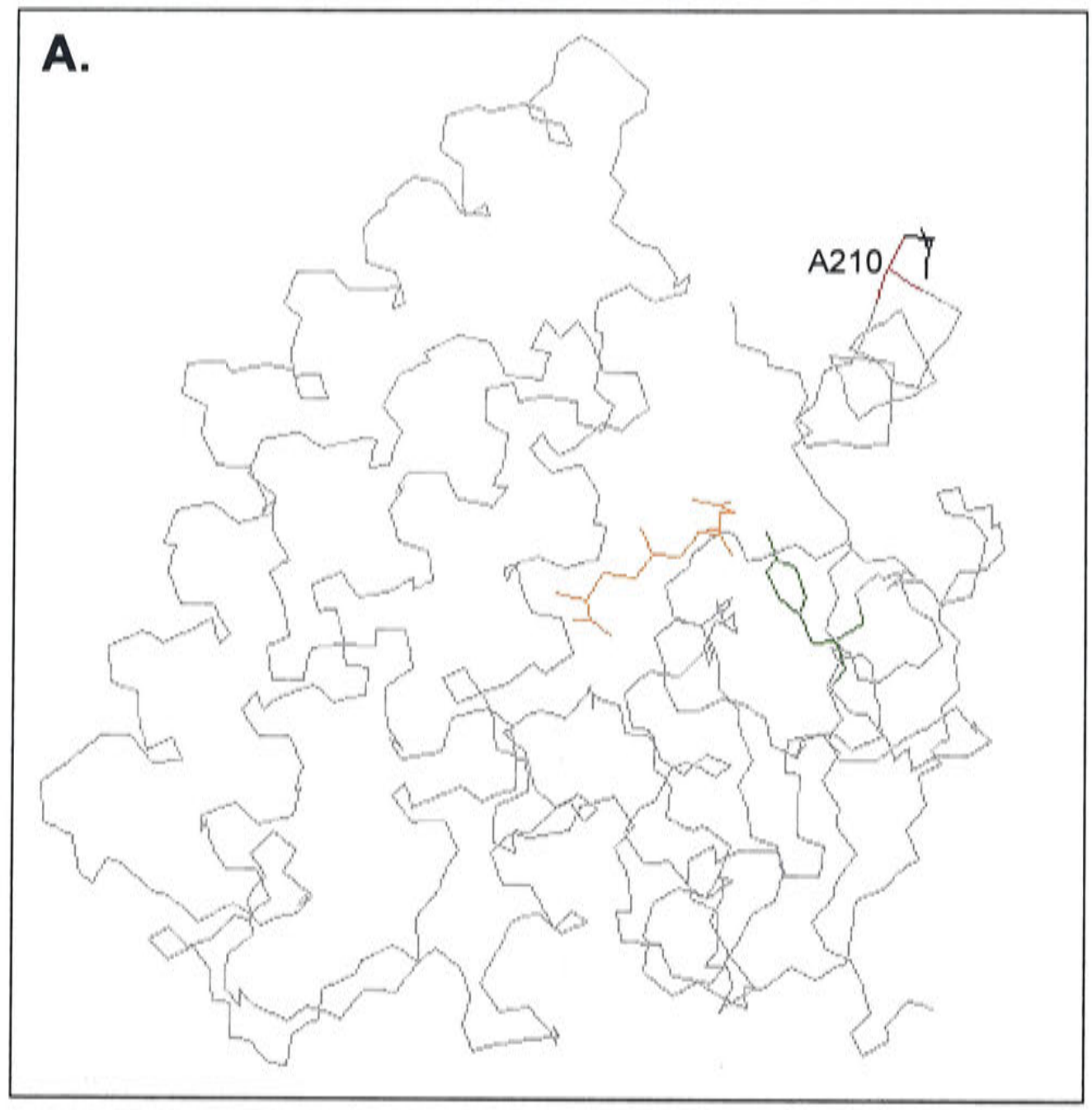

B.
Q113
T112/S112
F111
L109
L108
P110

Figure 3.9 – Carbon backbone of the GSTA2 monomer encoding p.E210 overlaid with the protein encoding p.A210. GSTA2 encoding the more common amino acid p.E210 (depicted in black) is coloured grey; GSTA2 encoding the less common amino acid p.S110 (depicted in red) is coloured blue. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.Y9 residue is coloured green. (A). Substitution of the labelled p.A210 residue for the p.E210 residue does not alter the overall structure of the GSTA2 protein. (B). Magnification of residue 210 and neighbouring residues showing the different H-bond patterns. H-bonds common to both variants are coloured dark blue; the H-bond lost with the substitution of p.A210 for p.E210 is shown in pink.


A.
A210

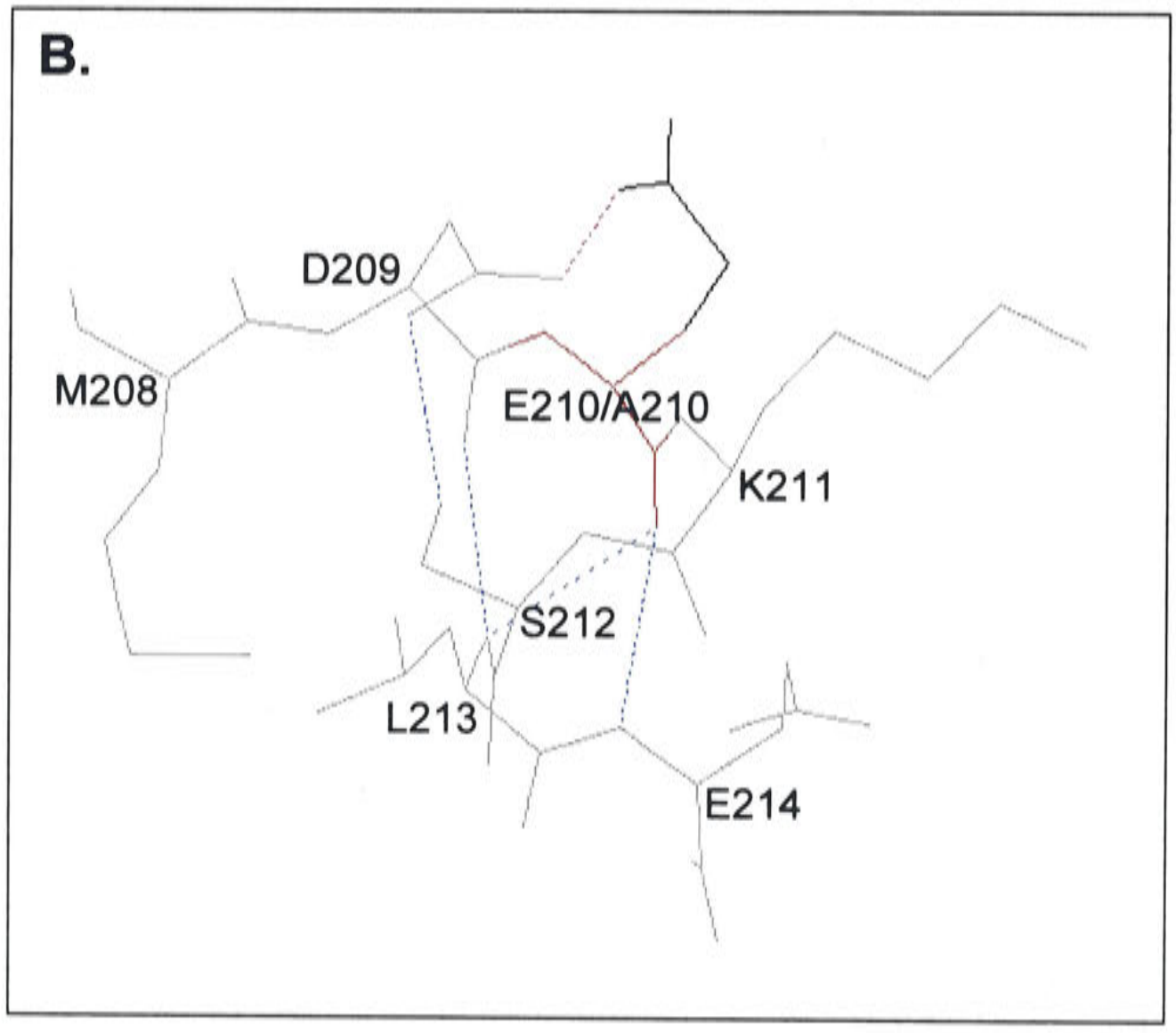

B.
D209
M208
E210/A210
K211
S212
L213
E214

Table 3.8 – Specific activities of recombinant GSTA2-2 variants towards various substrates.

Substrate	GSTA2A [p.P110;S112;E210]	GSTA2B [p.P110;S112;A210]	GSTA2C [p.P110;T112;E210]
		µmol/min/mg	
Δ^5-Androsten-3,17-dione[a]	0.22 ± 0.0075	0.21 ± 0.063	0.16 ± 0.048
1-chloro-2,4-dinitrobenzene[b]	21.47 ± 2.32	23.00 ± 3.48	26.69 ± 3.45
7-chloro-4-nitrobenz-2-oxa-1,3-diazole[b]	33.00 ± 2.40	35.36 ± 3.96	37.73 ± 3.66
Cumene hydroperoxide[b]	17.21 ± 1.86	21.72 ± 3.03	22.76 ± 2.18

[a] Results for CDNB, NBD-Cl and CuOOH are the mean ± S.E. for nine determinations.

[b] Results for Δ^5-AD are the mean ± S.D. for three determinations.

significant differences in activity for the p.T112 (GSTA2C) and p.A210 (GSTA2B) variants as compared to the GSTA2A protein. Large, significant differences were seen between the p.S110 variant and the GSTA2A protein (Table 3.9) with a ten-fold decrease in activity with Δ^5-AD ($p=0.04$), a two-fold decrease with CDNB ($p=0.004$), and a four-fold decrease with CuOOH ($p<0.0001$). A 6.2-fold increase was observed with *p*NPA ($p<0.0001$). No significant differences in activity between GSTA2A and GSTA2E were observed with the substrates DCNB, EA or NBD-Cl. No changes in expression levels were observed.

One point to note is the difference in specific activities measured for GSTA2A shown in Tables 3.8 and 3.9. This is due to the fact that the experiment presented in Table 3.8 was performed at a different time to that presented in Table 3.9 hence different preparations of enzyme and reagents were used. This emphasises that if enzyme activities are to be validly compared, the enzymatic assays must be performed at the same time and under the same conditions.

3.3.2.1.3 KINETIC STUDIES

Due to the large effect the p.P110S polymorphism has on the structure and function of the GSTA2A protein, kinetic studies were carried out to determine how this variant altered the catalytic efficiency of the GSTA2-2 protein. Kinetic parameters for CDNB were measured and Table 3.10 shows that the catalytic efficiency (K_{cat}/K_m) of the GSTA2E variant is only around 50% of the GSTA2A protein. This difference appears to be largely due to the difference in $K_{cat.}$ Unfortunately, although kinetic analyses were undertaken using *p*NPA as a substrate, the kinetic parameters could not be measured due to inconsistencies in the assay.

3.3.2.2 GSTA3-3

The p.L71I polymorphism in GSTA3-3 is a conservative change – both leucine and isoleucine are nonpolar residues of a similar size. It is unlikely that this change will

Table 3.9 – Specific activities of recombinant GSTA2-2 variants towards various substrates.

Substrate	**GSTA2A [p.P110;S112;E210]**	**GSTA2E [p.S110;S112;E210]**
	µmol/min/mg	
Δ^5-Androsten-3,17-dione	0.20 ± 0.094	0.015 ± 0.0072
1-chloro-2,4-dinitrobenzene	35.14 ± 2.71	16.82 ± 4.45
7-chloro-4-nitrobenz-2-oxa-1,3-diazole	93.27 ± 4.50	91.17 ± 1.48
Cumene hydroperoxide	14.72 ± 0.71	3.61 ± 0.41
1,2-dichloro-nitrobenzene	0.14 ± 0.01	0.15 ± 0.009
Ethacrynic Acid	0.048 ± 0.024	0.065 ± 0.029
p-nitrophenylacetate	0.14 ± 0.032	0.87 ± 0.017

The data represented are the mean ± S.D. for three determinations.

Table 3.10 – Kinetic parameters of recombinant GSTA2-2 proteins with the substrate CDNB.

Protein	**K_m^{CDNB} (mM)**	**K_m^{GSH} (mM)**	**K_{cat} (sec^{-1})**	**K_{cat}/K_m^{CDNB} ($s^{-1}mM^{-1}$)**	**K_{cat}/K_m^{GSH} ($s^{-1}mM^{-1}$)**
GSTA2A [p.P110;S112;E210]	0.65 ± 0.035	0.19 ± 0.040	34.15 ± 2.15	52.34	178.68
GSTA2E [p.S110;S112;E210]	0.45 ± 0.11	0.23 ± 0.064	14.62 ± 1.06	32.43	64.89

affect the structure or the function of the GSTA3-3 protein, as it is not situated in either the G-site or H-site of the protein (Figure 3.6). This alteration was unable to be characterised enzymatically, as the transcript of GSTA3-3 missing exon 3 (667 bp), a known alternative transcript, was initially cloned, and the full-length transcript (719 bp) that was subsequently cloned could not be expressed. Difficulties expressing GSTA3-3, especially at high levels, has been previously reported [Johansson & Mannervik, 2002].

3.4 Discussion

The Alpha class GSTs are involved in a diverse range of roles including steroid biosynthesis and providing protection against alkylating agents and the products of lipid peroxidation. Functional polymorphisms in the GST Alpha class genes can therefore be expected to have wide-ranging physiological consequences, altering susceptibility to disease and drug response. Discovery of novel polymorphisms in these genes will expand our understanding of the underlying mechanisms governing the roles these genes play.

The use of database-mining programs such as the BLAST alignment tool to search the EST database and the SNP Finder program to search UniGene clusters enabled detection of polymorphisms in the *GSTA1* and *GSTA2* genes (Table 3.1). Lack of variation detected in the *GSTA3* and *GSTA4* genes may reflect a true absence of polymorphism, or be due to poor representation of cDNA sequences encoding these genes in the two respective databases. This was particularly relevant to the *GSTA3* gene, which was represented by a mere 20 cDNAs in the EST database and eight in the UniGene cluster, as opposed to the 136 (EST) and 99 (UniGene) cDNAs for *GSTA1*, and 46 (EST) and 33 (UniGene) cDNAs for *GSTA2*. Low representation is generally attributed to genes that are neither highly expressed nor found in multiple tissues. This is likely to explain the low representation of *GSTA3* in the EST database since it is only expressed in a limited number of tissues (adrenal gland, ovary, placenta and testis) [Johansson & Mannervik, 2001a]. In contrast, GSTA1-1 and GSTA2-2 are highly expressed in the liver, and at lower levels in a series of other tissues [Strange *et al.*, 1984; Rowe *et al.*, 1997]. Similarly, although GSTA4-4 was represented by 95 cDNAs in the EST database and is widely expressed [Desmots *et al.*, 2001; Morel *et al.*, 2002], only eight to 16 partial cDNAs covered a particular section of the coding region, and only one cDNA was found in the UniGene cluster. The likelihood of detecting polymorphisms in these sequence databases increases if the gene of interest is represented by a greater number of full-length cDNA sequences from different cDNA libraries generated from a diverse pool of sources. The continual expansion of these sequence databases will enable the detection of more polymorphic variants in genes that are not currently well represented.

Databases dedicated to the collection of SNPs spanning the entire gene sequence from the 5' UTR through to the 3' UTR and derived from numerous sources were also analysed for GST Alpha gene submissions. Again, there was a distinct lack of coding region variants detected in the *GSTA3* and *GSTA4* genes (Table 3.2). *GSTA3* was not well represented in these databases, although non-coding region polymorphisms were present in all but EGPSNPs, the CGAP lists and HGBASE (data not presented). *GSTA4* was better represented and was found in all but the EGPSNPs database, although again, most of the submitted information pertained to the non-coding regions (data not presented). This implies that perhaps these two genes are not highly polymorphic. However, this is not entirely true as two polymorphisms in the *GSTA4* gene, not detected by any method used in this study, were detected by independent groups [Latsoudis *et al.*, 2000; Iida *et al.*, 2001]. It is likely that as more sequencing information is gathered, a greater number of polymorphisms will be detected in these genes.

Unlike the BLAST alignment tool and the SNP Finder programs that were used to search different sequence databases, whereby it was possible to verify polymorphisms by resequencing cDNA clones or by accessing the available sequence trace data, the polymorphisms submitted to the SNP databases cannot be easily confirmed, as the information is generally obtained through sequencing genomic DNA from anonymous groups of individuals. In some instances, such as when viewing CGAP submissions and some refseq data, access to electropherograms is provided through the SNP Finder program. Other databases, such as the GeneSNPs and EGPSNPs, have determined frequency data for some polymorphisms. Although the SNP databases generally try to confirm the validity of the polymorphisms in multiple individuals and/or multiple reactions through resequencing efforts, the simplest way to determine the validity of these listings for the purpose of this thesis was to perform small-scale population screens, although this probably limits the number of polymorphisms confirmed to those that are common. To counter the possibility that some polymorphisms may occur specifically in particular ethnic groups, the small-scale screening included representatives from the world's major ethnic subgroups.

A particular problem encountered when searching these databases, and when using UniGene clusters, is that sequence data from multiple isoforms is often confused and

therefore submitted as one gene, giving incorrect estimations of the number of polymorphisms present in that particular gene. This was highlighted by the fact that *GSTA1*, one of the major liver detoxification enzymes, was only found in the GeneSNPs database. Further investigation revealed that *GSTA1* sequences had in fact been submitted to various databases however, these had been catalogued as *GSTA2* sequences. Hence the alignment of these apparently homogenous GSTA2 submissions with the *GSTA2* cDNA sequence gave the impression that there were many *GSTA2* polymorphisms, when in reality these "polymorphisms" represented the nucleotide differences distinguishing *GSTA1* from *GSTA2*. This confusion of genes is common, and as the incorrect information is presented, the need to be aware of sequence similarity within members of multi-gene families must be emphasised.

Despite the problems inherent with SNP detection methods, as discussed above, each was able to detect variants not identified by the other methods. Five variants that altered amino acid residues were identified and verified by re-sequencing in the *GSTA1* and *GSTA2* genes using the BLAST alignment tool and SNP Finder programs (Table 3.1), and seven novel coding region variants were detected in all four GST Alpha genes in the SNP databases (Table 3.2). Of these, only five out of the 12 possible variants were verified and found in the Australian European, Bantu African, Creole African and Southern Chinese populations. With regard to the K196N *GSTA2* variant however, the alteration was found at the end of a repetitive string of bases, implying it is probably a sequencing error. Usually, this would have been filtered out, but analysis was continued as this variation was identified in two I.M.A.G.E. Consortium (LLNL) cDNA clones. It is possible that the remaining seven variants represent rare alleles. Alternatively, as the ethnic origin of the cDNA in the EST and UniGene databases is unknown, there is a possibility that these variants may be associated with particular ethnic groups. The latter option provides the most likely explanation as to why the p.A90V and p.V149A *GSTA2* variants were not confirmed in the population studies presented here. These polymorphisms were found in the GeneSNPs (associated with UUGC) and EGPSNPs (associated with UWGC) databases, both of which contribute to the Environmental Gene Project. It is highly probable that SNPs detected by both centres are real, as these are initially discovered whilst resequencing genes of relevance to the project. Any SNPs so identified are further confirmed in a PDR cohort, which contains European American, African American, Mexican-American, Native American and Asian

American samples. GeneSNPs calculated a heterozygote frequency of 0.01% for the p.A90V variant, and EGPSNPs found that the p.V149A variant was quite common (T=81.91%, C=18.09%; T/T=30, C/T=17). It is possible that the p.A90V and p.V149A *GSTA2* variants have evolved in the Americas, which is why they were not observed in the ethnic groups studied here.

Of the four missense polymorphisms that were identified and confirmed in this study, both the *GSTA2* p.T112S and p.E210A polymorphisms have previously been noted as sequence discrepancies [Hayes *et al.*, 1989; Röhrdanz *et al.*, 1992], however allele frequencies and functional characteristics of these polymorphisms had not been investigated. The p.T112 variant of the p.T112S polymorphism was found to be relatively rare and only found in the heterozygous state in the Australian European and two African populations, although this allele was noticeably more frequent in the Bantu African population. In contrast, the p.E210A polymorphism was quite common. The p.A210 allele was represented in all ethnic groups studied, ranging from a modest 8% in the Australian European population to 42% in the Bantu African population. Allele distributions calculated by EGPSNPS found a similar distribution in their American populations (E=70.51%, A=29.49%; A/A=19, A/C=17, C/C=3) to that found in the Creole African population studied here (Table 3.7). Despite the good representation of the p.A210 variant in all populations studied here, this was never seen in combination with the p.T112 variant, consistent with observations reported by Coles and colleagues [2000].

The GSTA2 p.P110S polymorphism, previously identified by mass spectrometry, was also not subjected to further investigations [Wang *et al.*, 2000]. This was originally found in six out of 63 pancreas samples [Wang *et al.*, 2000], and in keeping with this reasonably rare occurrence in the pancreas, the p.S110 variant was found to be relatively uncommon in the four populations studied here, especially in the Bantu African population in which only two heterozygous individuals were identified, a finding similar to that calculated by EGPSNPs in the American populations (C=97.83%, T=2.17%; C/C=44, C/T=2). The p.S110 allele was better represented in the Creole African and Southern Chinese populations, in which a total of four homozygous individuals were detected. It appears that this relatively rare p.S110 variant is only observed in combination with the common p.S112 and p.E210 variants.

Genotyping results for the fourth missense polymorphism discovered in this study, the novel GSTA3-3 p.L71I polymorphism, revealed that the p.L71 variant was absent in all populations tested except the two African populations, in which it is found at a modest frequency of 5%-15%. This polymorphism was also discovered during the recent characterisation of the *GSTA3* gene, along with two alternative transcripts [Johansson & Mannervik, 2001a]. The alternative transcript missing exon 3 was cloned during the present study, preventing characterisation and comparison of each variant. In light of the role GSTA3-3 plays in progesterone and testosterone biosynthesis pathway [Johansson & Mannervik, 2001a], cloning of the full-length cDNA is being continued in order to determine the effect this polymorphism has on the GSTA3-3 protein and hence how it may affect steroid biosynthesis. However, as both isoleucine and leucine have similar properties, and residue 71 is located in neither of the active sites of the protein, it is unlikely that this will have any functional consequences.

Recombinant proteins encoding the three *GSTA2* missense polymorphisms detected in this study were subsequently tested for evidence of functional and structural alterations. In keeping with the predictions based on 3-D modelling, which demonstrated a lack of local and global conformational changes, no significant alteration in activity towards four typical alpha class substrates was observed when the recombinant GSTA2A [p.P110;S112;E210], GSTA2B [p.P110;S112;A210] and GSTA2C [p.P110;T112;E210] proteins were tested. Although residue 210 is located at the beginning of the flexible C-terminal α-9 helix (residues 210 to 220) that forms a cap over the active site [Sinning *et al.*, 1993] (Figure 3.5), the external orientation of the residues in this position suggest that they may have little influence upon function. Hence at this stage there is little evidence that the observed differences in allele and haplotype frequencies regarding these three alleles are directly due to functional effects.

In contrast, significant alterations in the specific activity of the GSTA2E variant [p.S110;S112;E210] towards several substrates were measured. The specific activity of the GSTA2E isoform was generally lower than that of the GSTA2A enzyme, yet an increase in specific activity was seen towards *p*NPA (Table 3.9). Significant changes can be expected, considering that p.P110 is located at the C-terminus of the α4 helix (Figure 3.5), which along with the N-terminal loop and α9 helix contributes to the

Alpha active site [Sinning *et al.*, 1993]. Proline residues, which are hydrophobic, contribute to the structural rigidity of a protein. Hence substitution of a proline in the active site by a hydrophilic serine residue can be expected to both reduce the rigidity of, and alter the topography of the active site. In support of the observed enzymatic changes, predictive modelling demonstrated that this polymorphism caused minor changes to the overall protein structure. In addition, one new H-bond to residues neighbouring those involved in GSH or substrate binding were created upon the substitution of p.S110 for p.P110 (Figure 3.7B). It can be envisaged that these combined effects could increase the flexibility of the active site, and through the ability of serine to form H-bonds with surrounding residues and incoming substrates, also increase the hydrophilicity of the active site. These effects could lower the affinity of the GSTA2E protein for hydrophobic substrates, alter the progress of substrates into the active site or even alter the optimal fit of a substrate in the active site through differences in electrostatic interactions or steric hindrance, all of which could potentially alter the enzymes catalytic abilities.

The reduced affinity of the GSTA2E protein for hydrophobic substrates was demonstrated not only by the decreased specific activities seen for the hydrophobic substrates CDNB, Δ^5-AD and CuOOH, but further shown by the decreased catalytic efficiency (K_{cat}/K_m) of GSTA2E for CDNB and GSH, largely brought about by a 2.3-fold reduction in K_{cat} compared to GSTA2A. This may reflect perturbations in product release, or in the ability of the molecule to allow substrates in and products out. Kinetic data measured for GSTA2A was comparable to that measured by Zhao and colleagues [1999]. Although kinetic parameters were not measured for Δ^5-AD and CuOOH due to time constraints, it is likely that the decreased specific activities observed are also due to inefficient binding. In contrast, the introduction of a serine residue enhanced the activity of GSTA2-2 for *p*NPA, possibly due to the greater flexibility of the active site allowing the substrate better and faster access, or the ability of serine to act as a strong nucleophile upon this polar substrate. Unfortunately the mechanism behind the increased specific activity of the GSTA2E enzyme for the *p*NPA substrate could not be tested accurately in this study, for although kinetic experiments were undertaken, the results obtained were inconsistent. No change in specific activity was observed with the hydrophilic substrates EA or NBD-Cl.

Considering the high peroxidase activity possessed by the GSTA2-2 family [Stockman *et al.*, 1987; Chow *et al.*, 1988; Zhao *et al.*, 1999] and the five-fold reduction in specific activity observed here with the GSTA2E variant towards CuOOH, it is possible that an increased susceptibility to disease associated with oxidative stress may be observed in individuals expressing this isoform, perhaps even at a younger age. Disease is most likely to be manifested in tissues in which the variant GSTA2-2 is expressed, namely the liver. The liver is a prime target, both because GSTA2-2 is one of the major liver proteins [Strange *et al.*, 1984; Rowe *et al.*, 1997] and it is at an increased risk, in comparison to other organs, of suffering oxidative stress due to the large number of ROS generated during drug and xenobiotic metabolism [Zhao *et al.*, 1999].

Approximately one fifth of each population studied, except the Bantu African population, was heterozygous for the p.S110 variant. So, although the effect of this polymorphism in these individuals would be dampened by the high activity p.P110 allele, increased susceptibility to the products of oxidative stress could still be observed. It is unlikely that this polymorphism will have a direct causal link to any particular disease, especially as disease caused by oxidative stress is complex, involves interactions between many genes, and may also be influenced by an individual's environment and lifestyle. In addition, polymorphisms already characterised in the GST genes have been shown to exert only a relatively weak influence on phenotype [Hayes & Strange, 2000; Strange *et al.*, 2000]. In order to better determine the exact role this polymorphism may have on diseases related to oxidative stress, particularly hepatic diseases, patient groups representing these different diseases should be screened. Ten of the 19 Australian European individuals heterozygous for the p.S110 allele did not express the *GSTM1* gene (data not presented), another major liver protein [Rowe *et al.*, 1997]. As the Alpha class GSTs are by default the major liver GSTs in *GSTM1*0* individuals, it may be of interest to determine whether any associations can be made between the GSTA2-2 p.S110 variant and the *GSTM1*0* allele.

The decreased peroxidase activity exhibited by the GSTA2E protein may also extend to the prostaglandin endoperoxidase conversion pathway, in which GSTA2-2 uses its GPx-II activity to catalyse the reduction and isomerisation of the prostaglandin PGH_2 to PGD_2 and to a lesser extent, $PGF_{2\alpha}$ [Burgess *et al.*, 1989]. A decrease in peroxidase

activity could potentially interfere with PGD_2 and $PGF_{2\alpha}$ synthesis and function. PGD_2, a major prostaglandin in the central nervous system [Ogorochi *et al.*, 1984], is believed to have neuromodulator or neurotransmitter properties with involvement in the induction of hypothermia [Ueno *et al.*, 1982] and sleep [Hayaishi, 1988]. $PGF_{2\alpha}$ blocks adipogenesis through the activation of mitogen activated protein kinase hence is believed to play a role in the prevention of obesity and diabetes [Reginato *et al.*, 1998]. Yet as other pathways exist for the synthesis of these prostaglandins [Burgess *et al.*, 1989; Kanaoka *et al.*, 1997], the p.S110 variant is not likely to have a profound effect on PGD_2 or $PGF_{2\alpha}$ related functions.

The Alpha class GSTs have a protective role in detoxification of alkylating anti-neoplastic drugs, in particular the nitrogen mustard drugs. However, there is growing evidence that over-expression of the Alpha class GSTs is involved in the development of resistance to these drugs in tumour cells [Lewis *et al.*, 1988; Hayes & Wolf, 1990; Tsuchida & Sato, 1992; Tew, 1994]. GSTA2-2 has been implicated in resistance to chlorambucil [Lewis *et al.*, 1988; Ciaccio *et al.*, 1991; Meyer *et al.*, 1992], the alkylating agent used in treating Hodgkin's and non-Hodgkin's lymphoma, chronic lymphocytic leukemia and ovarian cancer [Meyer *et al.*, 1992]. It is possible that the p.S110 variant may alter resistance to these hydrophobic drugs, as point mutations are one of many ways known to alter drug resistance [Hayes & Wolf, 1990].

When considering the multitude of effects polymorphisms may have on protein function, it is extremely important to take interindividual variation of protein expression into account. Various studies have demonstrated marked interindividual and gender variation of GSTA1-1 and GSTA2-2 expression, both within tissues and between tissues [Hayes *et al.*, 1989, Singhal *et al.* 1993, Mulder *et al.*, 1999; Coles *et al.*, 2000; Coles *et al.*, 2002]. It has been suggested that the variation in enzyme activity caused by this variation may in fact be greater than that caused by genetic polymorphisms [Coles *et al.*, 2000]. The recent discovery that GSTA2-2 expression is regulated by allelism in the *GSTA1* promoter [Coles *et al.*, 2001a] should also be considered, as it is possible that the effect of the p.S110 variant may be modulated to some degree by the *GSTA1* genotype.

Using a combination of database-mining methods, a moderate level of genetic polymorphism was identified in the human Alpha class GST genes. Although each method yielded results, this study has shown that identification of polymorphism should not be limited to one method or technique, especially since some genes are not as well represented as others in the various databases. Of the four missense polymorphisms verified in three ethnic populations: the p.P110S, p.T112S and p.E210A *GSTA2* polymorphisms and the p.L71I *GSTA3* polymorphism, functional and structural differences were observed in only one, the *GSTA2* p.P110S variant. Further characterisation is required to determine the physiological effects.

CHAPTER 4

THE MU CLASS GSTS

4.1 INTRODUCTION

The Mu class GSTs represents another diverse family of GSTs, possessing five isoforms with wide ranging functions. The Mu class GSTs are involved in the detoxification of epoxides, as well as prostaglandin synthesis and peroxidation of the oxidation products of the catecholamines cyclic *o*-quinones. Polymorphism in the Mu class genes was first suggested when it was noticed that GSTμ (now known as *GSTM1*), which is highly expressed in the human liver, was not detected in some individuals [Warholm *et al.*, 1980]. Further studies showed that this absence of expression was due to a null allele (*GSTM1*0*) [Board, 1981a; Strange *et al.*, 1984], caused by deletion of the *GSTM1* gene [Seidegård *et al.*, 1988; Xu *et al.*, 1998b]. This finding prompted interest in the detection and identification of other GST Mu class polymorphisms that may impact upon the detoxification abilities of these proteins.

Of the five GST Mu enzymes, GSTM1-1 shows the highest affinity for detoxifying epoxide metabolites such as styrene 7,8-oxide, trans stilbene-oxide and AFB_1 *exo*-8,9-epoxide [Warholm *et al.*, 1983; Hayes & Pulford, 1995; Johnson *et al.*, 1997], which are generated by the breakdown of PAHs and aflatoxin B_1 by Phase I enzymes. These mutagenic and oncogenic compounds are found in products of combustion, such as cigarette smoke, car exhaust fumes and burnt or mouldy foods. As mentioned above, the first Mu class polymorphism identified was the *GSTM1*0* allele. The frequency of this null allele varies considerably between populations, ranging from 21%-30% in Chilean and some African populations [Zhao *et al.*, 1994; Mukanganyama *et al.*, 1997; Masimirembwa *et al.*, 1998; Quiñones *et al.*, 1999; Johansson & Mannervik, 2001b], to 50% in the Caucasian population [Board, 1981a; Strange *et al.*, 1984; Seidegård *et al.*, 1988] and 100% of the Micronesian population [Board *et al.*, 1990]. In 1986 the proposal that *GSTM1*0* individuals were unable to metabolise epoxides [Board, 1981b; Warholm *et al*, 1983] was confirmed when an association between smokers homozygous for the *GSTM1*0* allele and an increased risk for lung cancer due to higher levels of PAH-adducts was reported [Seidegård *et al.*, 1986; Shields *et al.*, 1993]. A large number of studies have since tried to link the *GSTM1*0* allele, either singly or in combination with other alleles such as the *GSTP1* V105 and *CYP1A1* alleles [Hayashi *et al.*, 1992; Ryberg *et al.*, 1997], to a significant increase in susceptibility to disease

caused by epoxide exposure. Studies have concentrated especially on smoking related cancers such as lung and bladder cancers [Bell *et al.*, 1992; Ryberg *et al.*, 1994; Tang *et al.*, 1998; Salagovic *et al.*, 1999; Johns & Houlston, 2000; Aktas *et al.*, 2001; Engel *et al.*, 2002] in addition to stomach and colon cancers [Strange *et al.*, 1991; Zhong *et al.*, 1993b], with conflicting conclusions [Seidegård *et al.*, 1990; Zhong *et al.*, 1991; Daly *et al.*, 1993; Hirvonen *et al.*, 1993; Brockmöller *et al.*, 1998; Gertig *et al.*, 1998; Houlston, 1999; Welfare *et al.*, 1999; Benhamou *et al.*, 2002].

Paradoxically, the *GSTM1*0* allele has also been shown to confer protective effects. A two- to three-fold increase in relapse-free survival has been observed in children diagnosed with acute lymphoblastic leukemia (ALL) or acute myeloid leukemia (AML) carrying the *GSTM1*0* allele [Hall *et al.*, 1994; Chen *et al.*, 1997; Davies *et al.*, 2000; Stanulla *et al.*, 2000]. However, a direct association between *GSTM1* and other prognostic features of ALL has not been demonstrated [Chen *et al.*, 1997; Anderer *et al.*, 2000]. Also, protection against colorectal adenomas in *GSTM1*0* individuals by anti-carcinogenic isothiocyanates found in cruciferous vegetables has been recorded [Lin *et al.*, 1998].

In contrast to those individuals lacking a functional *GSTM1* gene, two Saudi Arabians have been discovered with a duplication of the *GSTM1* gene, which possesses ultra-rapid enzyme activity and thus enables rapid xenobiotic detoxification [McLellan *et al.*, 1997]. In addition, two functionally identical *GSTM1* alleles have been described: *GSTM1*A* and *GSTM1*B*. Although these differ by one amino acid, with *GSTM1*A* encoding a lysine at residue 173 [DeJong *et al.*, 1988] and *GSTM1*B* an asparagine [Seidegård *et al.*, 1988] and there is a charge change, they are functionally identical and are able to form active heterodimers with each other [Widersten *et al.*, 1991].

In addition to epoxide metabolism, *GSTM1* and *GSTM2* possess highly efficient peroxidase activity towards the cyclic *o*-quinones: aminochrome, dopachrome, noradrenochrome (*GSTM2*) and adrenochrome (*GSTM1*) [Baez *et al.*, 1997]. These GST Mu substrates are formed during the oxidation of the catecholamines dopamine, dopa, noradrenaline, and adrenaline respectively in a process that generates excessive numbers of ROS through redox cycling. This can lead to oxidative stress, toxicity, apoptosis and ultimately neurodegeneration and is believed to be involved in the

aetiology of Parkinson's Disease (PD) [Baez *et al.*, 1995]. This neurological disorder is characterised by the degeneration of dopaminergic neurons in the pars compacta of the substantia nigra of the human brain, a region found to express both GSTM1-1 and GSTM2-2 [Baez *et al.*, 1997; Ahmadi *et al.*, 2000]. Conjugation of dopamine and dopa to GSH, catalysed by GSTM2-2, has been shown to prevent the formation of their respective cyclic *o*-quinones and subsequent redox cycling [Dagnino-Subiabre *et al.*, 2000]. It has therefore been suggested that GSTM1-1 and GSTM2-2 may possess a neuroprotective role [Baez *et al.*, 1997]. Although no association has been made between PD and the *GSTM1*0* allele [Stroombergen & Waring, 1999; Ahmadi *et al.*, 2000; Harada *et al.*, 2001], the onset of PD was prolonged in patients with a viable *GSTM1* gene [Ahmadi *et al.*, 2000] and males with the *GSTM1*0* allele who had been exposed to various toxins were discovered to be more susceptible to PD [Stroombergen & Waring, 1999]. To date, the only *GSTM2* polymorphism known is a silent mutation involving asparagine at residue 74 [Vorachek *et al.*, 1991; Ross & Board, 1993]. Considering that GSTM2-2 is one of the more highly expressed Mu class GSTs in the brain [Rowe *et al.*, 1997], it can be postulated that other *GSTM2* polymorphisms resulting in functional change could influence the aetiology of PD.

GSTM3-3 and GSTM2-2 have been shown to possess cytosolic prostaglandin E synthase activity [Ogorochi *et al.*, 1987; Beuckmann *et al.*, 2000]. Employing this activity both enzymes are believed to be involved in the GSH-dependent isomerisation of PGH_2 to PGE_2 in the brain. PGE_2 is responsible for mediating several neurophysiological roles in the central and peripheral nervous systems. As both GSTM2-2 and GSTM3-3 are the most highly expressed Mu class GSTs in the brain [Rowe *et al.*, 1997], polymorphism in the *GSTM3* or *GSTM2* genes may disrupt these functions. Three polymorphisms have been identified in the *GSTM3* gene, however two remain uncharacterised: the glycine to tryptophan substitution at residue 147, believed to have no functional effect [Pearson *et al.*, 1993; Ross & Board, 1993], and the valine to isoleucine substitution at residue 224 [Emahazion *et al.*, 1999; Iida *et al.*, 2001]. A third polymorphism consisting of two alleles, *GSTM3*A* and *GSTM3*B*, is distinguished by a three base pair deletion in intron 6 [Inskip *et al.*, 1995]. *GSTM3*B* has been associated with altered expression levels of GSTM3-3 due to the presence of a negative transcription factor YY1 (yin yang 1) recognition motif in *GSTM3*B* created by the deletion of these three bases. YY1 is able to intragenically regulate gene

expression through activation or repression of transcription [Hariharan *et al.*, 1991; Shi *et al.*, 1991]. A significant association has been found between the *GSTM3*B* and *GSTM1*A* alleles, and it is believed that susceptibility to different diseases may be determined by the combined epistatic effects of the different *GSTM1* and *GSTM3* alleles [Inskip *et al.*, 1995; Yengi *et al.*, 1996]. Various studies have since tried to associate interactions between *GSTM1* alleles and expression of GSTM3-3 with different cancers and have generally found that the *GSTM3*B*B* genotype in combination with the *GSTM1* gene has a protective effect. A decreased risk to multiple basal cell carcinoma [Yengi *et al.*, 1996], laryngeal squamous cell carcinoma and other upper aerodigestive tract cancers has been observed in Caucasians expressing the *GSTM1*A* or *GSTM1*B* and *GSTM3*B* alleles [Jahnke *et al.*, 1996; Matthias *et al.*, 1998]. However, there is conflicting evidence of the role the *GSTM3* alleles play in oral cancers [Jourenkova-Mironova *et al.*, 1999a; Jourenkova-Mironova *et al.*, 1999b] and ethnic differences have been observed [Worrall *et al.*, 1998; Park *et al.*, 2000].

Polymorphism has also been recorded in the *GSTM4* and *GSTM5* genes. The GSTM4-4 isoenzyme has extremely low activity with the typical range of GST Mu substrates and a specific substrate has not been found, although a weak correlation between GSTM4-4 and PGF synthase has been made [Beuckmann *et al.*, 2000]. A silent polymorphism involving phenylalanine at residue 178 has been recognised [Comstock *et al.*, 1993; Pearson *et al.*, 1993; Ross & Board, 1993; Zhong *et al.*, 1993a] and analysis of the literature has revealed the presence of an unrecognised and uncharacterised V212M polymorphism [Comstock *et al.*, 1993; Ross & Board, 1993; Zhong *et al.*, 1993a]. In addition, two alternatively spliced GSTM4-4 transcripts have been identified, and both are believed to encode catalytically non-functional proteins [Ross & Board, 1993]. One transcript is distinguished by an 82 bp deletion, which abolishes exon 4 and leads to a premature stop codon 36 amino acids downstream. In the second transcript exon 8 is missing. Instead, intronic sequence is read through from the end of exon 7, creating an additional six codons and an inframe stop codon. Relatively little has been reported about *GSTM5*, although an uncharacterised *Hind*III polymorphism has been detected [DeJong *et al.*, 1991; Pearson *et al.*, 1993].

This chapter describes the use of various database mining techniques to facilitate the discovery of novel Mu class polymorphisms. The distribution of the confirmed

polymorphisms was determined in three ethnic populations and the recombinant proteins were characterised enzymatically to determine whether the function or the structure of the protein was altered by the introduction of the polymorphism.

4.2 MATERIALS AND METHODS

Database analysis and standard PCR and recombinant methods used in this chapter are described in Chapter 2. Techniques unique to this chapter are outlined below.

4.2.1 DATABASE SCREENING

Potential polymorphisms in the five Mu class GSTs were identified and verified using the database mining methods described in §2.3, and are listed in Tables 4.1 and 4.2.

4.2.2 PCR/RFLP ANALYSIS

Small-scale population studies were performed by specifically amplifying potentially polymorphic exons of the *GSTM1-GSTM5* genes from genomic DNA using PCR as described in §2.5.6.2. Oligonucleotides and conditions specific for each PCR are detailed in Table 4.3. The presence of candidate polymorphisms in the PCR products was determined by RFLP analysis (§2.5.4.2) using the appropriate restriction enzymes (Table 4.4). Two of the variants tested altered restriction endonuclease sites. For the remaining variants, primers were designed containing a partial restriction endonuclease site, which was completed by either the absence or the presence of the variant base [Kangsadalampai *et al.*, 1998]. Up to 100 samples from the Australian European, Bantu African, Creole African and Southern Chinese population groups were amplified by PCR, to allow allele and haplotype frequency determination for polymorphisms confirmed by the small-scale population screens.

Table 4.1 – GST Mu variants detected in the EST database by the BLAST alignment tool and in the UniGene database by the SNP Finder program.

Gene	Nucleotide	Residue	Exon	Program	ESTs with alterations	cDNA libraries represented	Sequence Confirmed[a,b]	Confirmed in population studies
GSTM1	c.518C>G	p.N173K[c]	7	BLAST	8	5	Yes-AA227608	Yes
	c.527C>T	p.D176D	7	BLAST	7	4	Yes-AA227608	Yes
GSTM2	c.222T>C	p.N74N[d]	4	BLAST	6	5	N.S.	N.S.
	c.226A>G	p.E89G	5	SNP Finder	1	1	Yes-0.05	No
	c.614C>A	p.P205Q	7	SNP Finder	2	2	No-0.01	N.S.
	c.628A>T	p.T210S	8	SNP Finder	7	5	Yes-0.44	N.S.
GSTM3	c.371A>C	p.H124P	5	SNP Finder	3	3	No-0.02	N.S.
	c.384G>C	p.K128N	6	SNP Finder	1	1	Yes-0.44	No
	c.439T>G	p.W147G[d]	6	BLAST	40	21	N.S.	Yes
	c.549A>G	p.P183P	7	SNP Finder	2	1	Yes-0.66	N.S.
	c.572G>T	p.R191L	7	SNP Finder	1	1	Yes-0.83	No
	c.670G>A	p.V224I	8	BLAST	13	8	Yes-AI870867	Yes
	c.670A>G	p.I224V	8	SNP Finder	18	13	Yes-1.0	
GSTM4	c.165A>G	p.L55L[d]	3	BLAST	7	5	N.S.	N.S.
GSTM5	c.103A>C	p.M35L[d]	2	BLAST	7	5	N.S.	No

[a] The cDNA clones sequenced to confirm the presence of potential polymorphisms identified using the BLAST alignment are identified.

[b] Polymorphisms detected using the SNPfinder program are allocated a likelihood score. As the number approaches 1.0, there is a greater likelihood that this polymorphism will be real. Sequences were confirmed by viewing the electopherograms made available at the site.

[c] This is the known *GSTM1*A* to *GSTM1*B* polymorphism [De Jong *et al.*, 1988; Seidegård *et al.*, 1988].

[d] These polymorphisms were detected as sequence discrepancies i.e. the BLAST query sequence was the only sequence with the original residue. However, evidence of the *GSTM2* p.N74N and *GSTM3* p.W147G polymorphisms are present in the literature: the *GSTM2* clone of Ross & Board [1993] had the c.222C variation, and that of Vorachek *et al.* [1991] the c.222T variation; the *GSTM3* clone of Pearson *et al.* [1993] had the c.439G variation and that of Campbell *et al.* [1990] the c.439T variation. The *GSTM5* p.M35L polymorphism was later studied in small-scale population studies.

N.S.: Not studied

Table 4.2 – GST Mu polymorphisms detected in the SNP databases.

Gene	Nucleotide	Residue	Exon	Database	Able to Confirm[a]	Confirmed in population studies
GSTM1	c.84T>C	p.Y28Y	2	dbSNP, refseq, GeneSNPs, SNPper	Unable	N.S.
	c.306C>T	p.N102N	5	refseq	Unable	N.S.
	c.519C>G	p.N173K[b]	7	[f]CGAPV (0.38), dbSNP, SNPper	Yes	Yes
	c.528C>T	p.D176D[b]	7	[f]CGAPV (0.99), dbSNP, GeneSNPs, SNPper	Yes	Yes
	c.562T>A	p.F188I	7	[f]CGAPV (0.11)	Yes	No
	c.628T>A	p.S210T[c]	8	[f]CGAPV (0.99), dbSNP, refseq, GeneSNPs, HGBASE	Yes	N.S.
GSTM2	c.120T>C	p.D40D	3	dbSNP, refseq, SNPper	Unable	N.S.
	c.223T>C	p.N74N[b]	4	dbSNP, refseq, GeneSNPs, EGPSNPs, SNPper	Unable	N.S.
	c.357T>A	p.D119E	5	refseq	Unable	No
	c.372A>G	p.K124K	6	refseq	Unable	N.S.
	c.382T>C	p.L128L	6	refseq, GeneSNPs	Unable	N.S.
	c.385C>G	p.Q129E	6	refseq, GeneSNPs	Unable	No
	c.389C>A	p.A130E	6	dbSNP, refseq, GeneSNPs	Unable	No
	c.401T>A	p.M134K	6	dbSNP, refseq, GeneSNPs	Unable	No
	c.405G>A	p.L135L	6	dbSNP, refseq, GeneSNPs	Unable	N.S.
	c.559C>G	p.R187G	7	refseq	Unable	No
	c.628A>T	p.T210S[b,c]	8	dbSNP, refseq, GeneSNPs, EGPSNPs, HGBASE	Unable	N.S.
GSTM3	c.222C>A	p.I74I	4	HGBASE	Unable	N.S.
	c.371A>C	p.H124P[b]	5	CGAPV (0.02)	Unable	N.S.

	c.384G>C	p.K128N[b]	6	CGAPV (0.44), refseq, HGBASE	Yes	No
	c.384G>T	p.K128N	6	CGAPV (0.44)	No	N.S.
	c.549A>G	p.P183P[b]	7	CGAPV (0.66), dbSNP, HGBASE	Yes	N.S.
	c.572G>T	p.R191L[b]	7	CGAPV (0.83), refseq, HGBASE	Yes	No
	c.670G>A	p.V224I[b]	8	CGAPV (1.0), dbSNP, refseq, GeneSNPs, HGBASE	Yes	Yes
GSTM4	c.534T>C	p.F178F[d]	7	dbSNP, refseq, GeneSNPs, SNPper	Unable	N.S.
	c.634G>A	p.V212M[e]	8	dbSNP, refseq, GeneSNPs, SNPper	Unable	No
	c.657A>G	p.X219X	8	dbSNP, GeneSNPs, SNPper	Unable	N.S.
GSTM5	c.103C>C	p.M35L[b]	2	refseq	Unable	No

[a] Some sequences were unable to be confirmed directly through the databases used therefore small-scale population screens were used to confirm these polymorphisms. Those detected through the validated CGAP lists could be confirmed through viewing existing electropherograms using the SNP Finder program.

[b] These polymorphisms were also detected by the BLAST alignment tool and the SNP Finder program.

[c] This polymorphism was detected in both *GSTM1* and *GSTM2*. It is unlikely to be a real polymorphism, rather represents both genes.

[d] This silent *GSTM4* polymorphism has been noted. Comstock *et al.* [1993], Ross & Board [1993] and Zhong *et al.* [1993a] detected the c.534C variant; Pearson *et al.* [1993] detected the c.534T variant.

[e] Analysis of the literature has revealed the presence of this polymorphism. Comstock *et al.* [1993] and Ross & Board [1993] isolated the p.V212 variant; Zhong *et al.* [1993a] isolated the p.M212 variant.

[f] These polymorphisms were detected in sequences submitted as *GSTM4* in the validated CGAP lists although the sequences represent *GSTM1*, due to the fact that the *GSTM1*B* sequence template X08020 was used as the query sequence. The sequences that were submitted under *GSTM1* in the candidate CGAP lists were of unknown identity and are listed as being located on chromosome X or chromosome 19 (AK001936) in the SNP index. The *GSTM1* sequences submitted into the CGAP validated lists (CGAPV) were also of unknown identity and W95015, which is 100% identical to AK001936, was used as the query template.

Table 4.3 – Oligonucleotides used for *GSTM1-5* exon amplification.

Primer	Exon	Primer sequence 5' to 3'	Annealing Conditions	Product Length
GSTM1				
M1Ex71F	7	GATCACTTTTGTAGATTTTCTCGT	54°C, 18 sec	228 bp
M1Ex7R		GAAACCAGTACTCAATACATGCG		
M1Ex7F	7	CTCCACCGTATATTTGACACCAA	61°C, 18 sec	72 bp
M1Ex7R2		CTCAAAGCGGGAGATGAAGTCC		
GSTM2				
M2Ex5F	5	GCTGAGAGTGAATCTGCTTTAGCA		128 bp
M2Ex5R		ATCTGGGTCATAGCAGAGTTTGGC	57°C, 20 sec	
M2Ex5R2		GACACAAACTCTGCAGGGAAAGG	54°C, 18 sec	205 bp
M2Ex6F	6	CTGGGGCCATGCACAAAAC	57°C, 18 sec	227 bp
M2Ex6R		TAGGAAGTATGACAAATCTGTG		
M2Ex7F	7	TTTGTCATACTTCCTATATTATG	53°C, 16 sec	202 bp
M2Ex7R		AGACATAAAGAGAAAGGAGGC		
M1/2Ex8F	8	TCCCCTTACTAGGTATTTA	51°C, 18 sec	635 bp
M1/2Ex8R		GAGTGAAGAGGGACAATG		
GSTM3				
M3Ex6F	6	CCTGTGCCCTGATTAACT	53°C, 18 sec	226 bp
M3Ex6R		CTTAGGTCTGAGGAGTAG		
[a]M3Ex6AflII		CCTCTCAGGAAAAACT**T**AA	51°C, 18 sec	
M3Ex7F	7	CACCTATGATATCTTGGATCAG	58°C, 18 sec	205 bp
M3Ex7R		CAACCATTGATCCCATTAGGC		
[b]M3Ex8NlaIII	8	CCAGTGGGGCAACAAGC**A**T	55°C, 18 sec	133 bp
M3Ex8R		GGACACCAGTAACATAAGTG		
GSTM4				
M4Ex8F	8	CCCCCTTAGTAGCTATTTG	55°C, 18 sec	667 bp
M4EX8A		GAGACAGAACCAGTCTACG		
GSTM5				
M5Ex2F	2	ACTCAAGCTATGTGGAAAAGCAGTAC	61°C, 18 sec	343 bp
M5Ex2R		TCCCACATCCCAGGGTTCAGGGAG		

[a] A modification of g→t (underlined) was introduced to create an *Afl*II site in the M3Exon 6 amplified DNA. This primer was used with the M3Ex6R primer.

[b] A modification of gcc→gca (underlined) was introduced to created an *Nla*III site in the M3Exon8 amplified DNA.

Table 4.4 – Restriction Endonucleases used to detect polymorphisms in *GSTM1-5*.

Variation	Exon	Endonuclease Site	Fragment Size (bp)		
			Wild-type	Heterozygous	Polymorphic
GSTM1					
p.D176D	7	*Fok*I	119,109	119,109,83,26	119,83,26
p.F188I	7	*Msl*I	228	228,118,110	118,110
GSTM2					
p.E89G	5	*Sau*3AI	128	128,92,36	92,36
p.D199E	5	*Apo*I	205	205,126,79	126,79
p.Q129E	6	*Hpy*CH4V	120,70,26,11	190,120,70,26,11	190,26,11
p.A130E	6	*Cac*8I	118,109	227,118,109	227
p.M134K	6	*Alu*I	96,64,44,23	96,64,58,44,23,6	96,58,44,23,6
p.R187G	7	*Hpa*II	202	202,159,43	159,43
GSTM3					
p.K128N	6	*Afl*II	155, 15	170, 155, 15	170
p.W147G	6	*Bst*XI	226	226,125,101	125, 101
p.R191L	7	*Eco*NI	205	205,119,86	119,86
p.V224I	8	*Nla*III	113,20	133,113,20	133
GSTM4					
p.V212M	8	*Fok*I/*Msc*I	323,249,95	323,249,171,95,78	323,171,95,78
GSTM5					
p.M35L	2	*Mwo*I	199,57,39,26,13,9	199,83,57,39,26,13,9	199,83,39,13,9

4.2.3 GSTM3-3 Recombinant Protein Work

4.2.3.1 Site-Directed Mutagenesis

The GSTM3-3 protein was expressed in *E. coli* in the plasmid pKK261 and was confirmed by cycle sequencing to be the GSTM3A protein (Table 4.5). Using the QuikChange Site-Directed Mutagenesis Kit (§2.6.3), the GSTM3B protein [p.G147;I224] was created from GSTM3A using the primer pair M3V224IA and M3V224IB (Table 4.6). The GSTM3C [p.W147;V224] and GSTM3D [p.W147;I224] proteins were created from the GSTM3A and GSTM3B proteins respectively using the primer pair GSTM3G147WA and GSTM3G147WB (Table 4.6). All mutations were confirmed by cycle sequencing (§2.5.5.1).

4.2.3.2 Protein Expression and Purification

Plasmids encoding the GSTM3A, GSTM3B, GSTM3C and GSTM3D recombinant proteins were transfected into *E.coli* strains XL1-Blue or TG1 (§2.6.2). To express these, 4 mL of an overnight culture was diluted 1:100 into LB supplemented with 100 µg/ml ampicillin. The 400 mL cultures were grown at 37°C to an OD_{600} of 0.6 and protein expression was induced with 0.2 mM IPTG overnight. The cells were harvested by centrifugation at 5 K for 10 minutes at 4°C (GSA-1500 rotor), resuspended in 30 mL resuspension buffer (50 mM Tris-HCl, pH 7.2, 50 mM EDTA, 15% glucose) and lysed by sonication (5 x 15 second bursts). Cellular debris was removed by centrifugation at 12 K for 20 minutes at 4°C (SS-34 rotor).

The proteins were all purified at 4°C by affinity chromatography. Cleared lysates were incubated for one hour on a rotary mixer with hexyl-glutathione agarose. The agarose was collected by centrifugation at 1000 rpm for 5 minutes and washed twice with 40 mL of 50 mM Tris-HCl, pH 7.2. The agarose was washed with a further 500 mL of 50 mM Tris-HCl, pH 7.2 on a scintered funnel, then transferred to a column and washed with a further 100 mL of 50 mM Tris-HCl, pH 7.2. The proteins were eluted from the hexyl-glutathione agarose with 5 mM pentyl glutathione in 50 mM Tris-HCl, pH 9.6.

Table 4.5 – Nucleotide and amino acid variations in GSTM3-3.

Haplotype	Nucleotide	
	c.439	**c.670**
GSTM3A	G	G
GSTM3B	G	A
GSTM3C	T	G
GSTM3D	T	A
	Residue	
	p.147	**p.224**
GSTM3A	Gly	Val
GSTM3B	Gly	Iso
GSTM3C	Trp	Val
GSTM3D	Trp	Iso

Table 4.6 – Mutagenesis primers used to create GSTM3-3 recombinant protein variants.

GSTM3 Haplotype	Primer Name	Mutagenesis Primer Sequences 5' to 3'
GSTM3B	M3V224IA	CCCACAGTGAAGAAGTTTCTACAGCCTGGC
	M3V224IB	GCCAGGCTGTAGAAACTTCTTCACTGTGGG
GSTM3C/D	GSTM3G147WA	CAATTCTCCATGTTTCTGTGGAAATTCTCATGGTTTG
	GSTM3G147WB	CAAACCATGAGAATTTCCACAGAAACATGGAGAATTG

Fractions containing recombinant GSTM3 were identified by 12.5% SDS-PAGE (§2.5.2.3). The strongest fractions were pooled and dialysed in 50 mM Tris-HCl, pH 7.2, 1 mM EDTA and 0.2 mM DTT. After an initial 4-5 hours of dialysis, the protein was transferred to fresh dialysis buffer and left overnight. Protein concentrations were determined using the Peterson method (§2.5.2.3).

4.2.4 GSTM3-3 Enzyme Characterisation

4.2.4.1 Structural Analysis

The GSTM3A [p.G147;V224], GSTM3B [p.G147;I224], GSTM3C [p.W147;V224] and GSTM3D [p.W147;I224] variants were built into the GSTM3 chain of the 3-D structure of GSTM2-3 (PDB file 1GTU) as described in §2.6.5.

4.2.4.2 Enzymatic Analysis

Glutathione *S*-transferase activity assays towards a range of GST substrates were measured as described in §2.6.4. Three separate preparations of GSTM3A and GSTM3B, and two separate preparations of GSTM3C and GSTM3D were prepared.

4.2.4.3 Steady State Kinetic Analysis

Steady state kinetics for the recombinant proteins GSTM3A, GSTM3B, GSTM3C and GSTM3D against the substrate CDNB were determined using a 5x5 experiment. Kinetic parameters for the CDNB reaction were determined by measuring the kinetics for five concentrations of CDNB (1000, 800, 600, 400 and 200 μM) against five concentrations of GSH (1000, 500, 200, 100 and 50 μM GSH). The steady state kinetic parameters were determined by fitting the data points to the Michaelis Menten equation using the MacCurve fit program. Hyperbolae were fitted to bring R^2 closest to 1 (≤ 1) and to minimise the SSE (≥ 0.0001).

4.3 RESULTS

4.3.1 DETECTION OF MU CLASS POLYMORPHISMS

4.3.1.1 DETECTION USING DATABASE ANALYSIS PROGRAMS

The BLAST alignment tool and the SNP Finder program were used to detect polymorphisms in the Mu class GST cDNA sequences located in the human EST database and UniGene clusters respectively. The EST database contained 86 *GSTM1*, 88 *GSTM2*, 100 *GSTM3*, 64 *GSTM4* and 80 *GSTM5* cDNAs. However, when searching dbEST with the *GSTM1* and *GSTM2* cDNA sequences, 65 of the 86 and 88 *GSTM1* and *GSTM2* ESTs respectively were identical, indicating that only 21 cDNAs were unique to *GSTM1* and 23 unique to *GSTM2*. The UniGene clusters contained 65 *GSTM2*, 69 *GSTM3*, five *GSTM4* and 13 *GSTM5* cDNA and mRNA sequences. From these two sequence databases, ten variants that altered amino acid residues were detected in the *GSTM1, GSTM2*, *GSTM3* and *GSTM5* genes (Table 4.1). Of these, six variants were confirmed upon resequencing a representative I.M.A.G.E. Consortium (LLNL) cDNA clone or viewing existing electropherograms: the p.N173K variant in *GSTM1*, the p.E89G and p.T210S variants in *GSTM2* and the p.K128N, p.R191L and p.V224I variants in *GSTM3*. Two of the variants, the *GSTM3* p.W147G and the *GSTM5* p.M35L variants, were not initially sequenced as these were believed to be sequence discrepancies, that is, the query sequence used to search the EST database was the only sequence encoding the original residue: p.W147 in the case of the *GSTM3* sequence, and p.M35 in the case of the *GSTM5* sequence. The p.W147G variant was later detected during simultaneous screening for the known *GSTM3* intronic polymorphism IVS6+22-24delAGG. The *GSTM5* p.M35L variant was later studied in small-scale population studies as it had also been submitted to the public access SNP databases (Table 4.2).

4.3.1.2 DETECTION USING SNP DATABASE ANALYSIS

A number of SNP databases were also analysed for GST Mu class submissions.

Multiple variants were found in these databases, however when limited to those that altered amino acid residues in the coding region, nine potentially novel variants were found: p.F188I and p.S210T in *GSTM1*, p.D119E, p.Q129E, p.A130E, p.M134K and p.R197G in *GSTM2*, p.K128N in *GSTM3* and p.V212M in *GSTM4* (Table 4.2). Although some variants could be confirmed by analysing electropherograms and available frequency data, small-scale population analysis was undertaken for confirmation on all variants identified. Seven variants that had been detected using the BLAST alignment tool and SNP Finder programs were also detected in these databases.

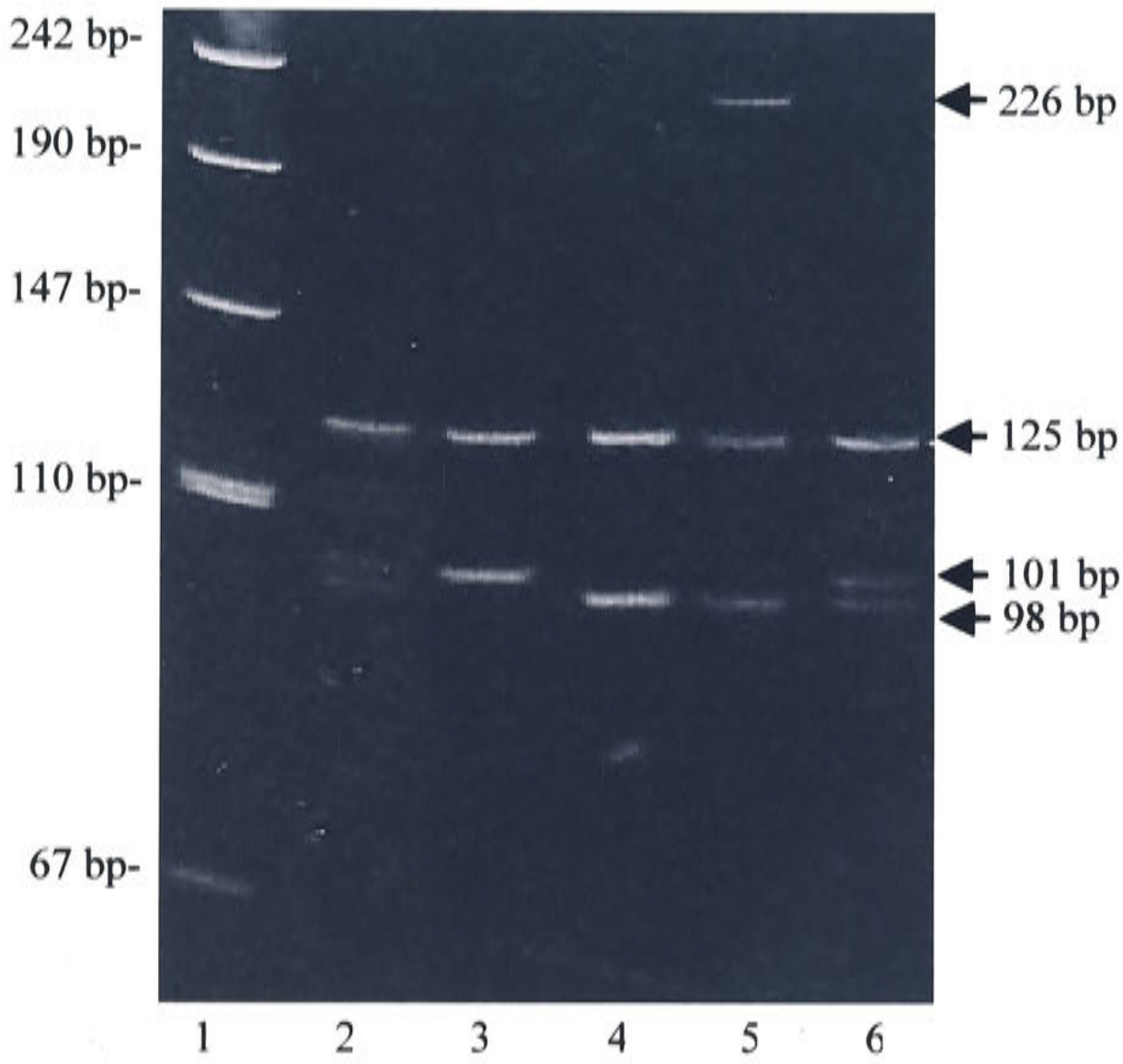
4.3.1.3 VARIANTS DISCARDED FROM FURTHER ANALYSIS

Three of the polymorphisms identified were not investigated further: the p.N173K and p.T210S *GSTM1* variants and the p.S210T *GSTM2* variant. The *GSTM1* p.N173K polymorphism has already been described and characterised [DeJong *et al.*, 1988; Seidegård *et al.*, 1988], however cDNA clones for this and a novel silent p.D176D polymorphism were further analysed as these two polymorphisms appeared to be linked. The two polymorphisms were confirmed by resequencing and found to occur in the Australian European population, but a link between the two polymorphisms was not found (data not shown). These results have not been presented as the true frequency of these alleles is confounded by the possibility that these samples are heterozygous for the *GSTM1*0* allele. The variation at residue 210 in GSTM1-1 and GSTM2-2 was not investigated further as it was assumed that this represented a database error. The nucleotide sequences of *GSTM1* and *GSTM2* are highly similar throughout intron 7, exon 8 and the 5' untranslated region. The substitution of c.628T (*GSTM1*) for c.628A (*GSTM2*), which changes p.S210 in GSTM1-1 to p.T210 in GSTM2-2, is the only difference observed in this region. It is likely that *GSTM1* submissions contain both *GSTM1* and *GSTM2* sequences. The same is true for *GSTM2* submissions.

4.3.1.4 Determination and Distribution of Allele Frequencies

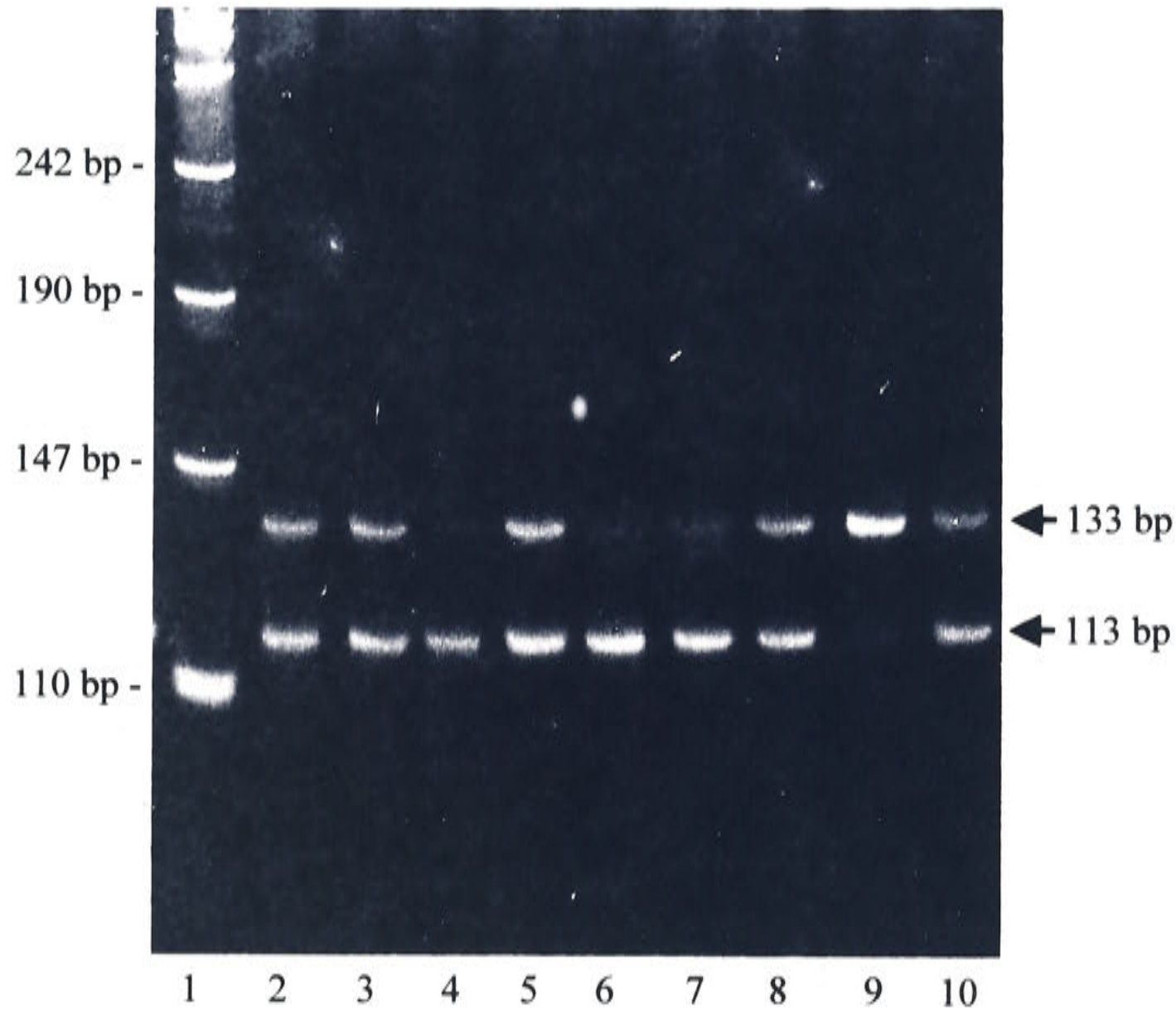
25 samples from three ethnic groups were tested for the presence of the 14 variants utilising PCR/RFLP assays to determine the nucleotide present at the potentially polymorphic site. Only two variants were ultimately identified and confirmed in the three populations: the *GSTM3* p.W147G (Figure 4.1) and *GSTM3* p.V224I (Figure 4.2) polymorphisms. Allele frequencies for these two polymorphisms were subsequently determined in up to 100 samples for the Australian European, Bantu African, Creole African and Southern Chinese populations (Table 4.7). The p.W147 variant of GSTM3-3 was found to be extremely rare, with only three heterozygous individuals detected in the Chinese population. No evidence for genotypic differences between the four populations was detected ($\chi^2_6 = 2.59$, $p>0.05$). The *Bst*XI restriction endonuclease used to genotype this polymorphism simultaneously genotyped the four populations for the known IVS6+22-24delAGG polymorphism (Figure 4.3). The *GSTM3*B* allele was found at a very high frequency of 80% in the Bantu African population compared with the 19% and 36% found in the Australian European and Creole African groups respectively. The *GSTM3*B* allele was absent in the Southern Chinese population. Significant genotypic differences were found between these four populations ($\chi^2_6 = 231.08$, $p<0.05$). The p.V224 variant, considered the common residue in GSTM3-3, was more frequent than the p.I224 variant in the Australian European (66%) and two African populations (67% and 89%). This situation was reversed in the Chinese population, where at 74%, the p.I224 variant was more common. Again, a significant genotypic difference was found between the four races ($\chi^2_6 = 134.66$, $p<0.05$). Further analysis demonstrated no evidence for genotypic differences between the Australian European and Creole African populations ($\chi^2_2 = 0.25$, $p>0.05$).

The p.W147G and p.V224I allelic variants of GSTM3-3 can be combined into four potential haplotypes as defined in Table 4.5. Only two of these were observed in the Australian European and two African populations, and three in the Chinese population (Table 4.7). This is based on the assumption that the combination of p.W147 with p.V224 is never observed in the Chinese population, an assumption based on the observation that two p.W147 heterozygotes were homozygous for the p.I224 variation, and a third p.W147 was heterozygous for the [p.V224;I224] polymorphism (Table 4.7).



Fragment (bp)	T/T	T/G	G/G
226	—	—	
125		—	—
101 and/or 98		—	—

Figure 4.1 - PCR/RFLP analysis of the *GSTM3* p.W147G variant. A 266 bp PCR product spanning exon 6 and part of intron 6 of the *GSTM3* gene was digested with *Bst*XI for nucleotide 439 determination. Lane 1 – pUC/*Hpa*II marker; Lanes 2-4, 6 - G/G homozygotes; Lane 5 - T/G heterozygote.



Fragment (bp)	G/G	G/A	A/A
133		—	—
113	—	—	
20	—	—	

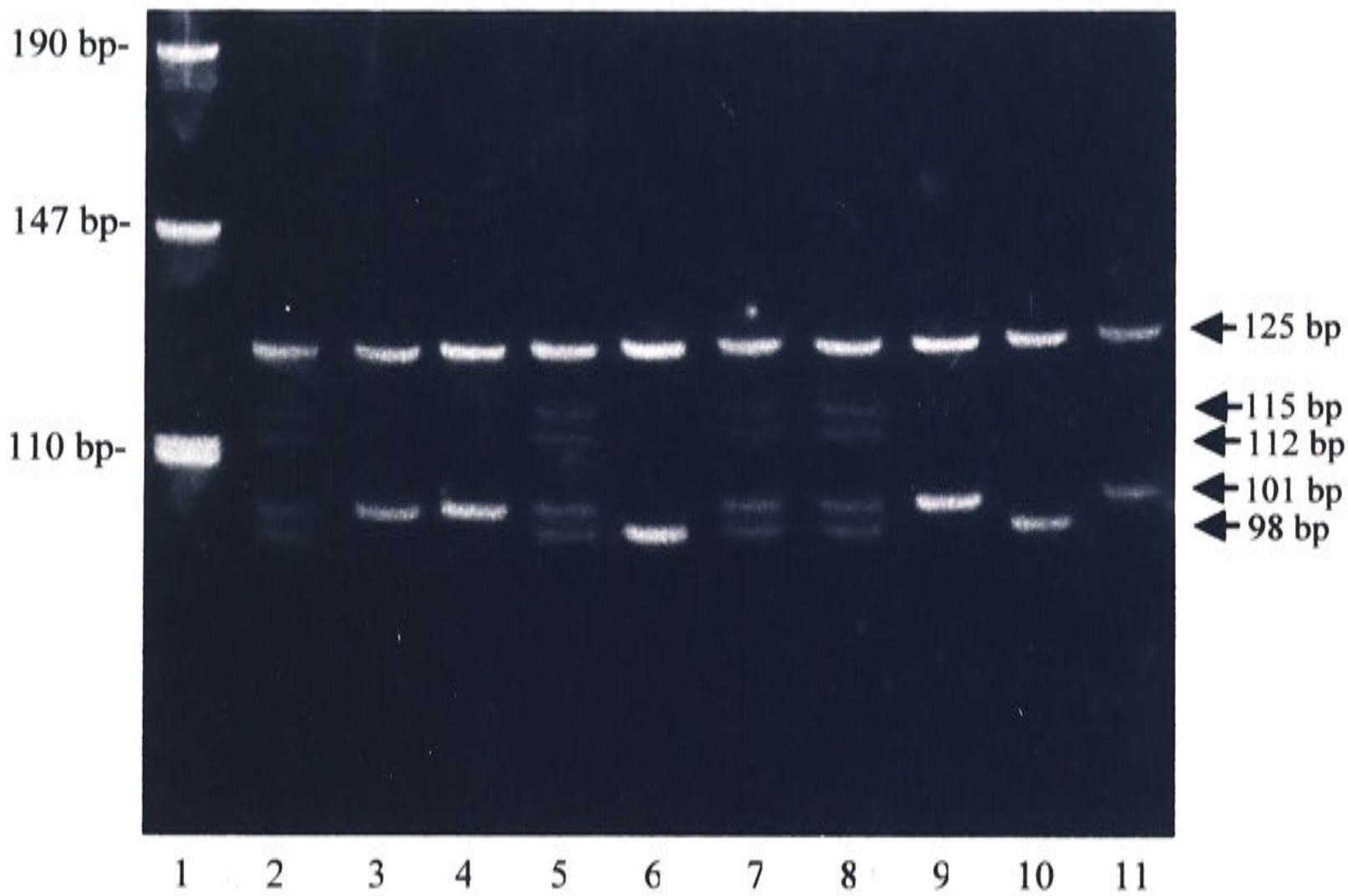
Figure 4.2 - PCR/RFLP analysis of the *GSTM3* p.V224I polymorphism. A 133 bp PCR product spanning exon 8 of the *GSTM3* gene was digested with *Nla*III for nucleotide 670 determination. Lane 1 - pUC/*Hpa*II marker; Lanes 2, 3, 5, 8, 10 - G/A heterozygotes; Lanes 4, 6, 7 - G/G homozygotes; Lane 9 - A/A heterozygote.

Table 4.7 – *GSTM3* allele frequencies in three ethnic groups.

Gene	Polymorphism	Population	n	Genotype			Allele Frequency	
GSTM3	p.W147G	Australian	100	T/T=0	T/G=0	G/G=100	p.W147=0	p.G147=1
		Bantu African	74	T/T=0	T/G=0	G/G=74	p.W147=0	p.G147=1
		Creole African	74	T/T=0	T/G=0	G/G=74	p.W147=0	p.G147=1
		Chinese	111	T/T=0	T/G=3	G/G=108	p.W147=0.01	p.G147=0.99
	IVS6+22-24delAGG	Australian	100	A*/A*=68	A*/B*=26	B*/B*=6	*A=0.81	B*=0.19
		Bantu African	61	A*/A*=2	A*/B*=20	B*/B*=39	*A=0.20	B*=0.80
		Creole African	74	A*/A*=29	A*/B*=36	B*/B*=9	*A=0.64	B*=0.36
		Chinese	111	A*/A*=111	A*/B*=0	B*/B*=0	*A=1	B*=0
	p.V224I	Australian	100	G/G=43	G/A=46	A/A=11	p.V224=0.66	p.I224=0.34
		Bantu African	70	G/G=54	G/A=16	A/A=0	p.V224=0.89	p.I224=0.11
		Creole African	69	G/G=30	G/A=33	A/A=6	p.V224=0.67	p.I224=0.33
		Chinese	93	G/G=4	G/A=40	A/A=94	p.V224=0.26	p.I224=0.74

Gene	Population	n	Haplotype		
			GSTM3A [p.G147;V224]	GSTM3B [p.G147;I224]	GSTM3D [p.W147;I224]
GSTM3	Australian	100	0.66	0.34	-
	Bantu African	57	0.89	0.11	-
	Creole African	68	0.68	0.32	-
	Chinese	92	0.255	0.728	0.016

Note: All polymorphisms are in agreement with the Hardy-Weinberg equilibrium



Fragment (bp)	*A/*A	*A/*B	*B/*B
125	—	—	—
101	—	—	
98		—	—

Figure 4.3 - PCR/RFLP analysis of the known *GSTM3* IVS6+22-24delAGG polymorphism. A 266 bp PCR product spanning exon 6 and part of intron 6 of the *GSTM3* gene was digested with *Bst*XI to determine *GSTM3*A* and *GSTM3*B* allele frequencies. Lane 1 – pUC/*Hpa*II marker; Lanes 2, 5, 7, 8 – *GSTM3*A*B* heterozygotes; Lanes 3, 4, 9, 11 – *GSTM3*A*A* homozygotes; Lanes 6, 10 – *GSTM3*B*B* homozygotes. It is likely that the bands appearing at approximately 112 and 115 bp in the *GSTM3*A*B* heterozygotes caused by the formation of heteroduplexes [Inskip *et al.*, 1995].

4.3.2 Enzyme Characterisation

4.3.2.1 Structural Analysis

To evaluate the potential effects of the p.W147G, and p.V224I variants of GSTM3-3, the substitutions were modelled into the GSTM3 chain of the 3-D structure of GSTM2-3 (Figure 4.4). Neither residues 147, located at the N-terminal end of the loop between the α5 and α6 helices, nor 224, located in the C-terminal extension unique to GSTM3-3, are located in the active site. However, predictive modelling demonstrates that both the p.W147G and p.V224I substitutions alone, and in combination, introduce varying degrees of conformational changes into the overall protein structure (Figures 4.5A, 4.6A and 4.7A). The p.W147G change is non-conservative, in which glycine, a small, uncharged polar residue is replaced with tryptophan, a bulky residue with a non-polar ring structure side chain. Although there is no evidence for altered H-bonding to residues immediately surrounding residue 147 (Figure 4.5B), alteration of the backbone surrounding the catalytically essential p.Y11 residue due to p.W147 creates one new H-bond between p.R15 and p.A18 (Figure 4.5C). In contrast, the p.V224I change is conservative – both valine and isoleucine are non-polar residues of similar size. Although this substitution would appear to have no effect on protein structure, backbone alterations in the immediate vicinity of residue 224 were observed, although H-bonding was not affected (Figure 4.6B). More importantly however, this substitution created a major change in the conformation of the catalytically essential p.Y11 residue, creating two new H-bonds: one from p.Y11 to p.L17, the other between p.R15 and p.A18 (Figure 4.6C). When seen in combination with the p.W147 variant, p.I224 adopted a different conformation to that observed with the p.G147 variant (Figures 4.6B and 4.7B). The conformation of p.W147 and p.Y11 residues as seen in Figures 4.5B and 4.5C, were not affected by the presence of the p.I224 variant.

4.3.2.2 Enzymatic Analysis

The activity of each of the four variants with a range of typical GST substrates was also examined, however activity was only observed with CDNB, CuOOH and *t*-nonenal

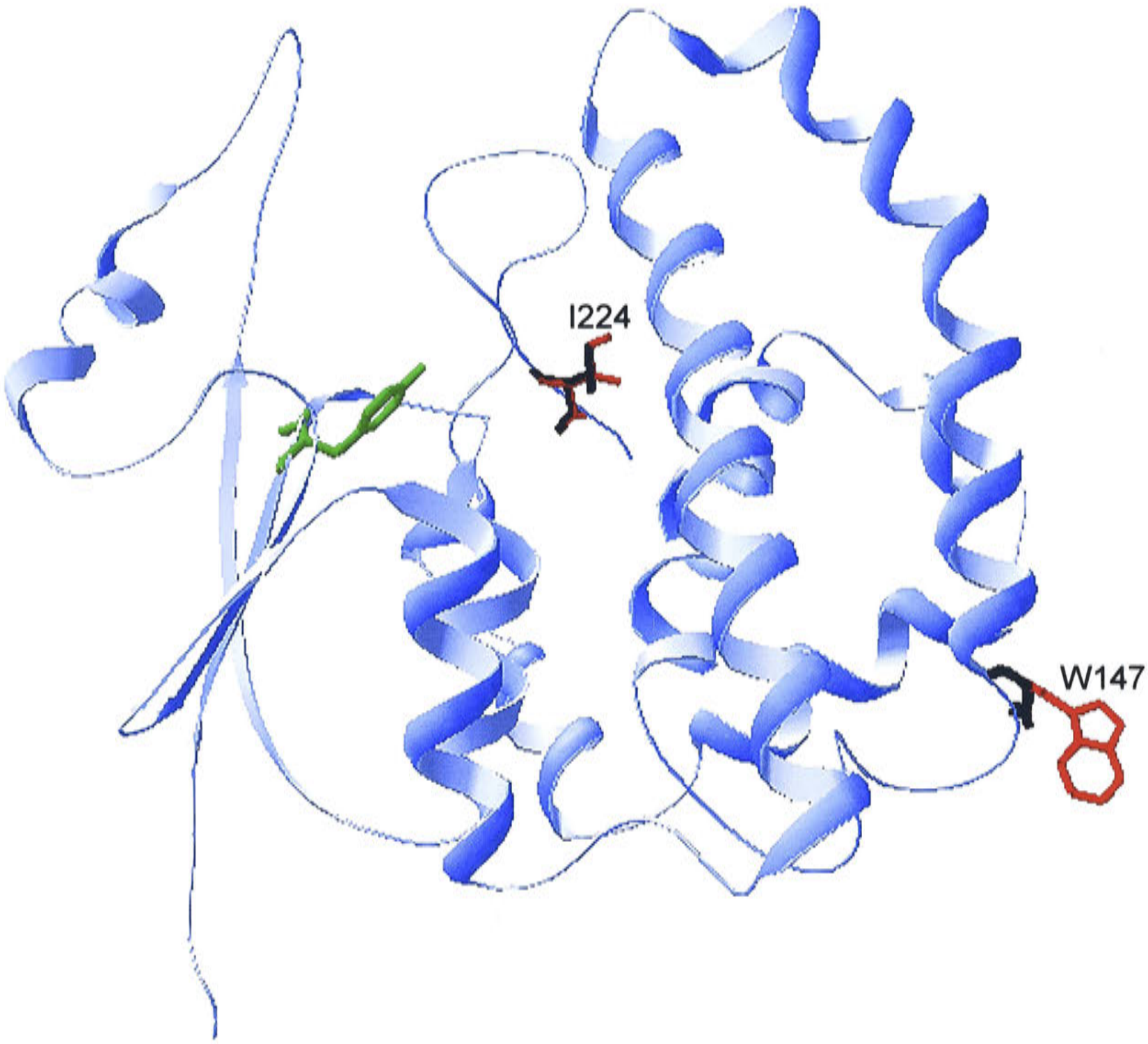
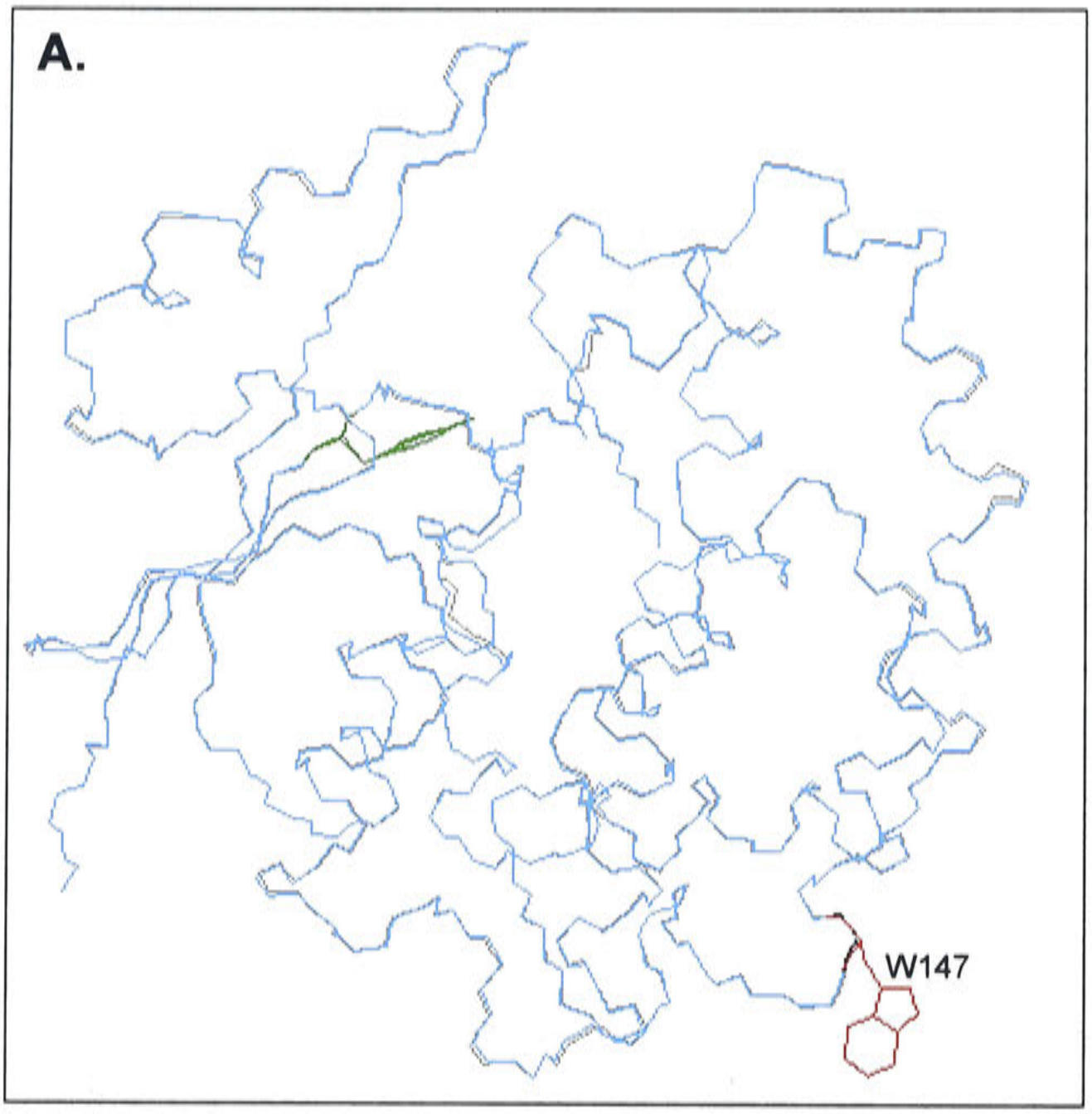
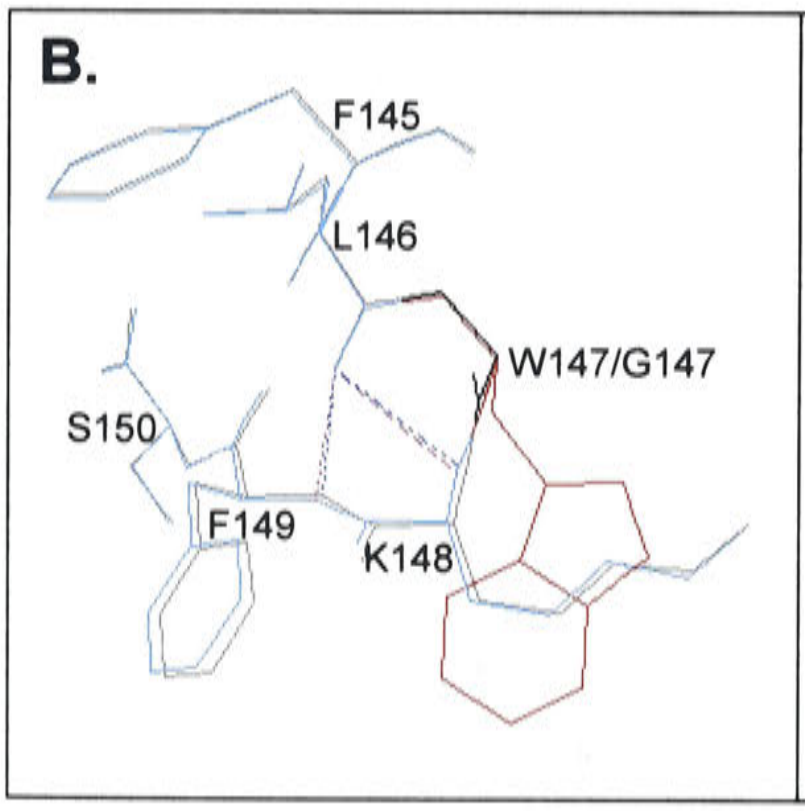


Figure 4.4 – Ribbon diagram of GSTM3 depicting the position of the polymorphic p.W147G and p.V224I substitutions. The most common amino acids, p.G147 and p.V224I, are shown in black. The least common amino acids, p.W147 and p.I224, are coloured red. The catalytically essential p.Y11 residue is coloured green.

Figure 4.5 – Carbon backbone of the GSTM3 monomer encoding p.G147 overlaid with the protein encoding p.W147. GSTM3 encoding the more common amino acid p.G147 (depicted in black) is coloured grey; GSTM3 encoding the less common amino acid p.W147 (depicted in red) is coloured blue. The catalytically essential p.Y11 residue is coloured green. (A). Substitution of the p.W147 residue for the p.G147 residue introduces minor changes to the overall structure of the GSTM3 protein. (B). Magnification of residue 147 and neighbouring residues. (C). Magnification of residue 11 and neighbouring residues. The H-bond introduced by the substitution of p.W147 for p.G147 is shown in pink.


A.
W147


B.
F145
L146
W147/G147
S150
F149
K148

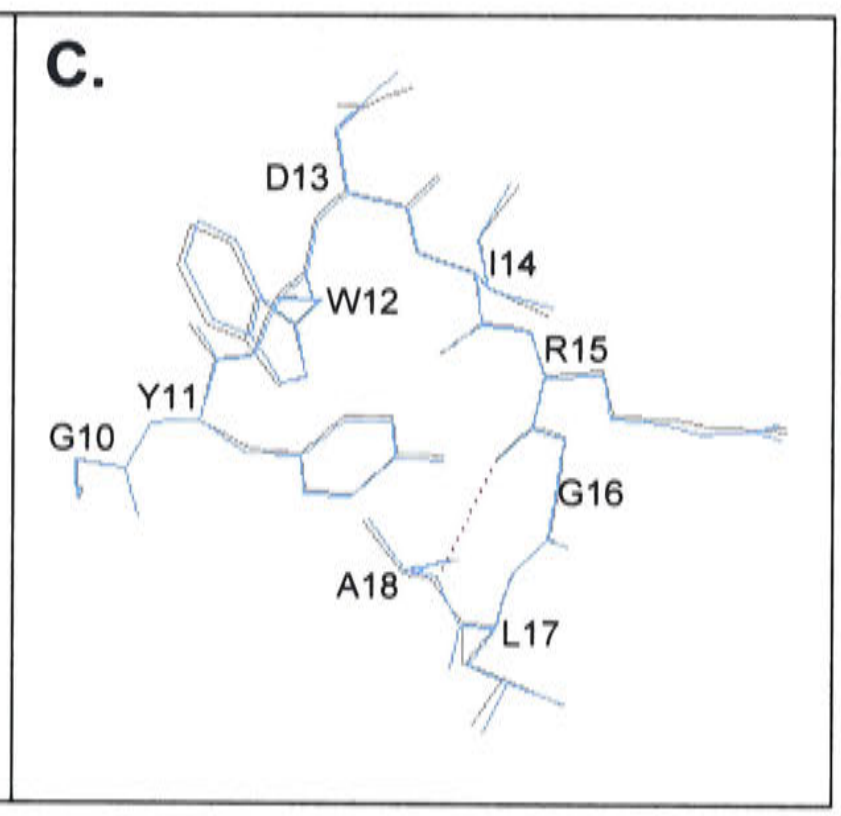
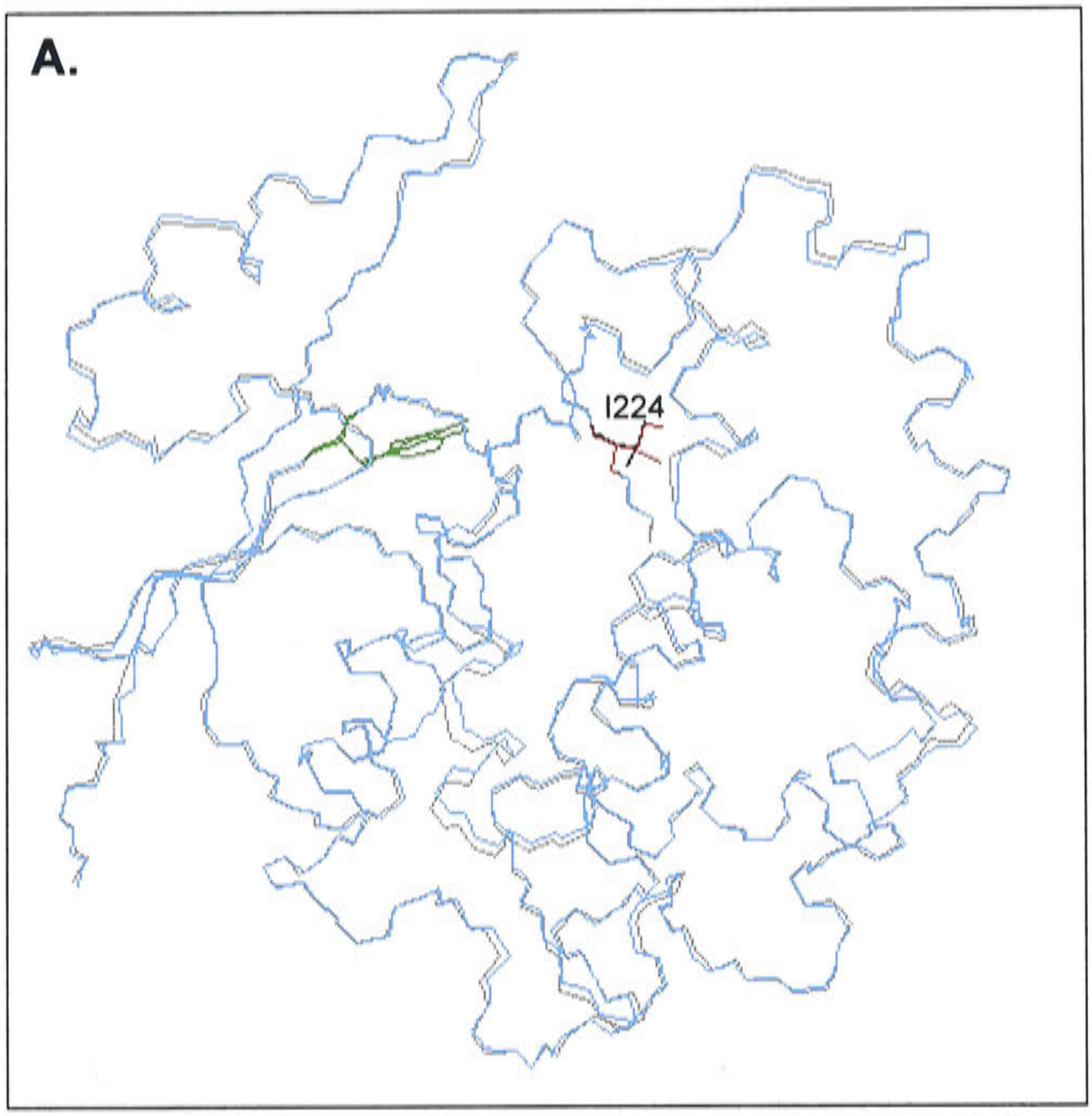
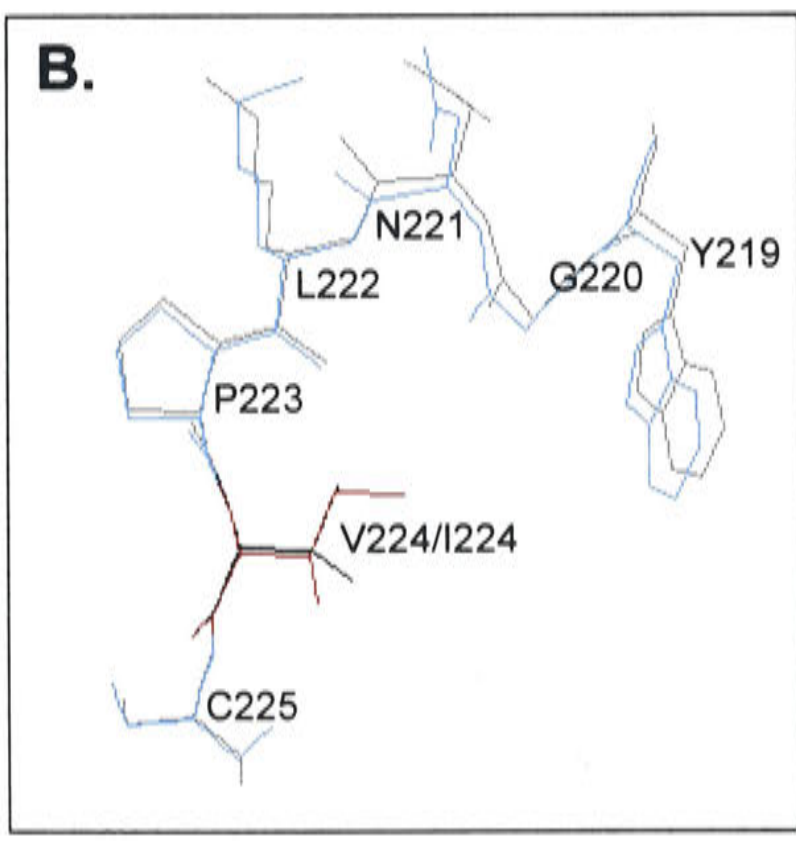

C.
D13
I14
W12
R15
Y11
G10
G16
A18
L17

Figure 4.6 – Carbon backbone of the GSTM3 monomer encoding p.V224 overlaid with the protein encoding p.I224. GSTM3 encoding the more common amino acid p.V224 (depicted in black) is coloured grey; GSTM3 encoding the less common amino acid p.I224 (depicted in red) is coloured blue. The catalytically essential p.Y11 residue is coloured green. (A). Substitution of the p.I224 residue for the p.V224 residue causes substantial changes to the overall structure of the GSTM3 protein. (B). Magnification of residue 224 and neighbouring residues. (C). Magnification of residue 11 and neighbouring residues. The H-bonds introduced by the substitution of p.I224 for p.V224 are shown in pink.


A.
1224


B.
N221
L222
G220
Y219
P223
V224/I224
C225

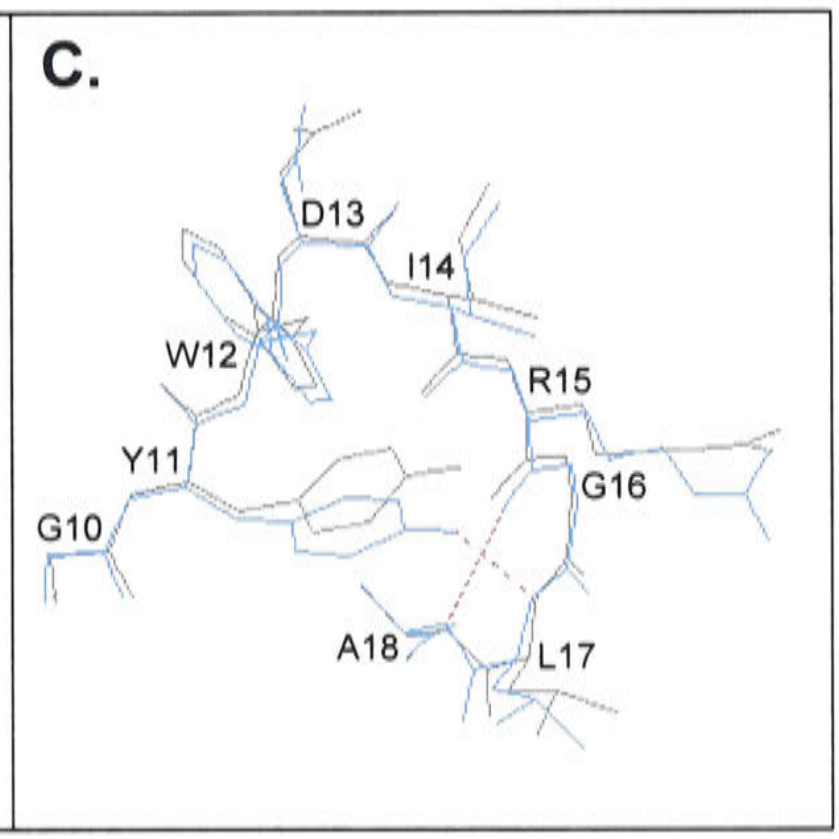
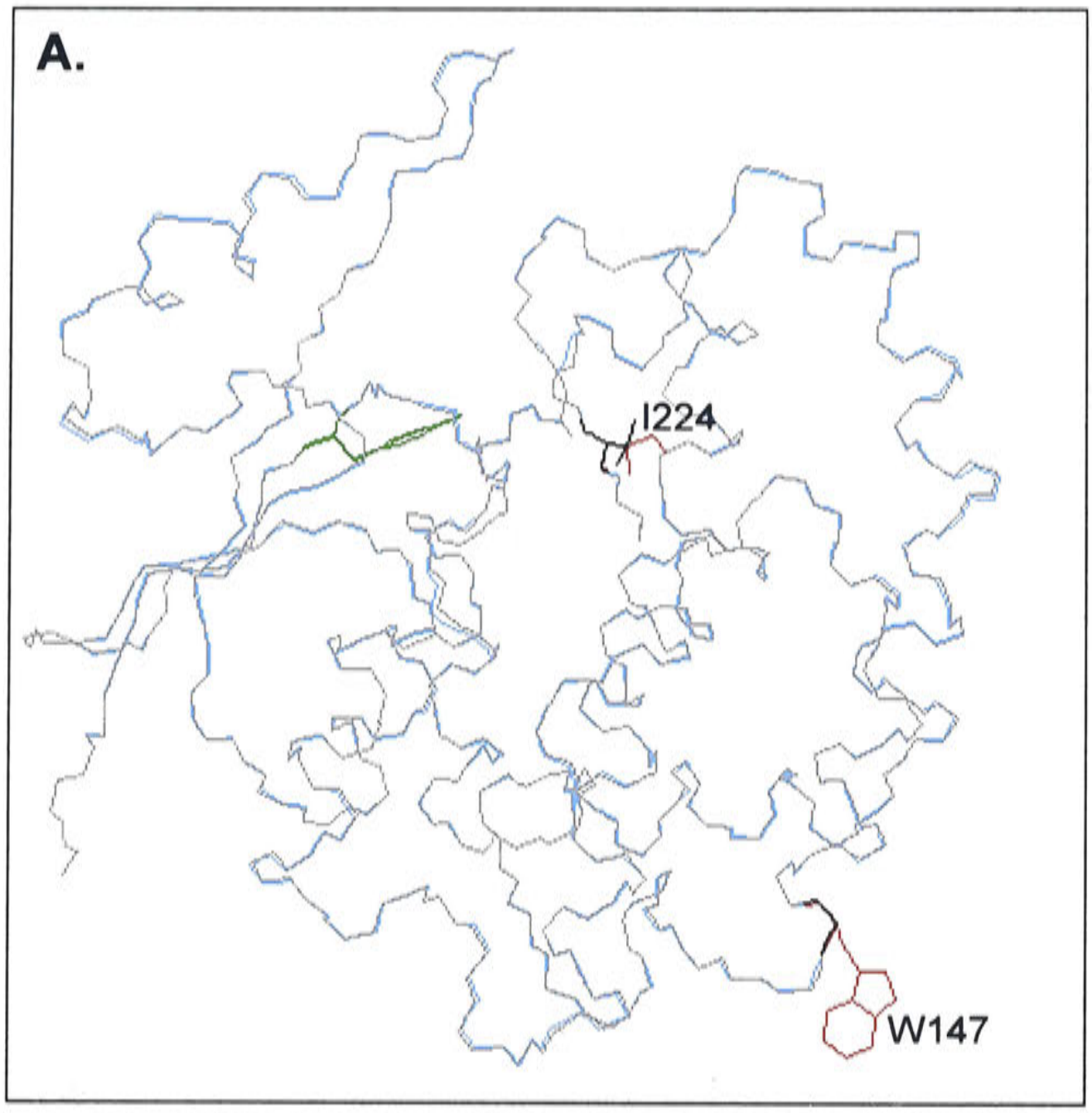
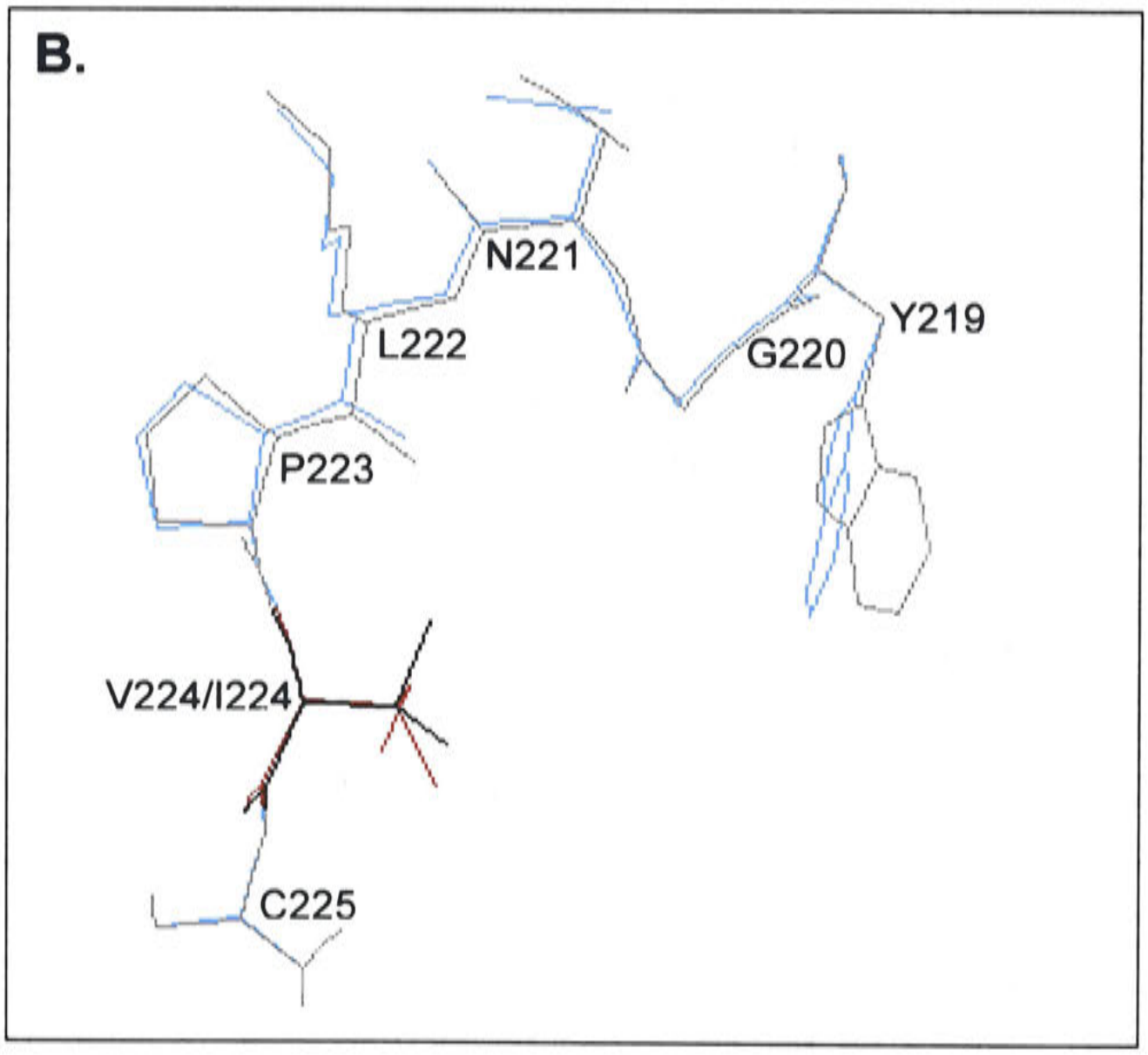

C.
D13
I14
W12
R15
Y11
G16
G10
A18
L17

Figure 4.7 – Carbon backbone of the GSTM3 monomer encoding [p.G147;V224] overlaid with the protein encoding [p.W147;I224]. GSTM3 encoding the more common amino acids [p.G147;V224] (depicted in black) is coloured grey; GSTM3 encoding the less common amino acids [p.W147;I224] (depicted in red) is coloured blue. The catalytically essential p.Y11 residue is coloured green. (A). Introduction of both the p.W147 and p.I224 residues into the GSTM3 protein causes minor changes to the overall structure. (B). Magnification of residue 224 and neighbouring residues.


A.
I224
W147


B.
N221
L222
Y219
G220
P223
V224/I224
C225

(Table 4.8). Introduction of p.W147 into the GSTM3-3 protein in combination with the p.V224 residue (GSTM3C) did not dramatically alter the specific activity of the GSTM3A protein [p.G147;V224], only causing a 1.3- to 1.5-fold decrease in specific activity towards the three substrates. This decrease was only significant for the substrate *t*-nonenal (p=0.009). Introduction of p.I224 into the GSTM3-3 protein containing the p.G147 residue (GSTM3B) significantly increased the specific activity towards CDNB by 2.8-fold (p=0.0015) and was responsible for a marginal, but non-significant, increase in the specific activities towards CuOOH and *t*-nonenal. In combination with the p.W147 residue, the p.I224 residue (GSTM3D) was responsible for an approximate two-fold decrease in specific activity towards CDNB (p=0.0129), a 1.8-fold decrease in CuOOH specific activity (p=0.0435) and a 1.7-fold decrease in specific activity with *t*-nonenal (p=0.0017), all of which were found to be significant. Differences in specific activities between the GSTM3B and GSTM3C isoforms were found to be significant for the substrates CDNB (p=0.0003) and *t*-nonenal (p=0.0162), as were those between the GSTM3C and GSTM3D isoforms (CDNB, p<0.0001; *t*-nonenal, p=0.0025). Significant changes were measured between the GSTM3B and GSTM3D isoforms for all substrates tested (CDNB, p=0.0002; CuOOH, p=0.0199; *t*-nonenal, p=0.003).

4.3.2.3 Kinetic Studies

Due to the large effects the p.W147G and p.V224I polymorphisms exerted on the function of the GSTM3A protein [p.G147;V224], kinetic studies were carried out to determine how these variants altered the catalytic efficiency of GSTM3-3. Kinetic parameters for CDNB were measured and are listed in Table 4.9. This analysis showed that substitution of p.W147 (GSTM3C) for p.G147 (GSTM3A) in GSTM3-3 encoding p.V224 caused a two-fold decrease in catalytic efficiency (K_{cat}/K_m) for CDNB, due to a two-fold decrease in K_{cat}. The K_m values for CDNB remained unchanged, but there was a slight 1.3-fold reduction in K_m for GSH. Introduction of the p.I224 variation into GSTM3-3 containing the p.G147 residue (GSTM3B) resulted in a 2.9-fold and 8.5-fold increase in catalytic efficiency for CDNB and GSH respectively, due to a three-fold increase in K_{cat}. The K_m values for CDNB remained unchanged, whilst that for GSH

Table 4.8 – Activities of recombinant GSTM3-3 variants towards various substrates.

Substrate	GSTM3A [p.G147;V224]	GSTM3B [p.G147;I224]	GSTM3C [p.W147;V224]	GSTM3D [p.W147;I224]
	(μmol/min/mg)			
1-chloro-2,4-dinitrobenzene	3.5 ± 0.91	9.76 ± 1.06	2.35 ± 0.01	1.24 ± 0.06
Cumene hydroperoxide	0.93 ± 0.27	1.02 ± 0.23	0.70 ± 0.24	0.38 ± 0.18
t-nonenal	1.72 ± 0.20	1.76 ± 0.26	1.16 ± 0.01	0.70 ± 0.12

The data represented are the mean ± S.D. for three determinations.

ANOVA analysis on Table 4.8

Significant differences in Enzyme activity were observed between the four different GSTM3 variants with the following substrates:

CDNB: $p = 0.0001$

CuOOH: $p = 0.0401$

t-N: $p = 0.0002$

Table 4.9 – Kinetic parameters of recombinant GSTM3-3 proteins with the substrate CDNB.

Protein	**K_m^{CDNB} (mM)**	**K_m^{GSH} (mM)**	**K_{cat} (sec^{-1})**	**K_{cat}/K_m^{CDNB} ($s^{-1}mM^{-1}$)**	**K_{cat}/K_m^{GSH} ($s^{-1}mM^{-1}$)**
GSTM3A [p.G147;V224]	1.50 ± 1.05	0.65 ± 0.11	4.85 ± 2.75	3.23	4.98
GSTM3B [p.G147:I224]	1.60 ± 0.89	0.35 ± 0.034	14.86 ± 2.96	9.26	42.07
GSTM3C [p.W147;V224]	1.56 ± 0.87	0.50 ± 0.06	2.36 ± 0.40	1.51	4.71
GSTM3D [p.W147;I224]	1.79 ± 0.77	0.43 ± 0.076	2.68 ± 0.43	1.50	6.23

was reduced approximately 50%. When the I224 variation was introduced into GSTM3-3 containing the p.W147 residue (GSTM3D), a 1.3-fold increase, compared to GSTM3A, in the catalytic efficiency for GSH was observed due to a 1.2-fold decrease in the K_m for GSH. A 50% decrease in the catalytic efficiency for CDNB due to a 1.2-fold increase in the K_m for CDNB, and a two-fold decrease in K_{cat} were also observed when GSTM3D was compared to GSTM3A.

4.4 DISCUSSION

The Mu class GSTs are believed to exert protective effects against numerous diseases ranging from epoxide induced cancers through to neurodegenerative disorders, in addition to maintaining PGE_2 regulated neurophysiological functions. Numerous studies have demonstrated weak associations between the well characterised *GSTM1*0* polymorphism with an increased risk to various epoxide induced cancers and have more recently found that the *GSTM3*B* intronic polymorphism confers a protective effect by reducing the risk of developing these cancers due to linkage with *GSTM1*A*. Polymorphism in other GST Mu isoenzymes could also possibly alter susceptibility to these diseases in addition to altering drug responses mediated by these GSTs.

Polymorphisms were detected in each of the Mu class GSTs by combining results obtained through searching the EST database with the BLAST alignment tool, and UniGene clusters using the SNP Finder program (Table 4.1). The BLAST alignment tool was able to detect variants in each of the five Mu class genes. Only one variant was detected in the *GSTM2*, *GSTM4* and *GSTM5* genes, however as these three genes were well represented in the EST database, detection of more variants could be expected. This lack of variation may be explained by the deficiency of full-length cDNA sequences. Only ten *GSTM2*, eight *GSTM4* and seven *GSTM5* partial cDNA sequences covered a particular section of the coding region of the respective genes. In contrast, the SNP Finder program detected many variants in the *GSTM2* and *GSTM3* genes, but not in the *GSTM4* and *GSTM5* genes despite their representation by five and 13 cDNA sequences respectively. The *GSTM1* UniGene cluster could not be downloaded or analysed due to unexplained problems with the database.

The lack of polymorphism detected by the SNP Finder program and BLAST alignment tool in the *GSTM4* and *GSTM5* genes, and to a lesser extent the *GSTM2* gene, may reflect a true absence of polymorphism. However the poor representation of full-length cDNA sequences encoding these genes in the two respective databases provides a more likely explanation. Under-representation of the *GSTM4* and *GSTM5* genes in particular may have been due to low expression levels. Although the GSTM4-4 and GSTM5-5 proteins are expressed in multiple tissues, neither is expressed at significant levels

[Comstock *et al.*, 1993; Takahashi *et al.*, 1993; Rowe *et al.*, 1997]. The other GST Mu isoenzymes are also expressed in multiple tissues to varying degrees [Warholm *et al.*, 1980; Laisney *et al.*, 1984; Beckett *et al.*, 1990; Campbell *et al.*, 1990; Takahashi *et al.*, 1993; Rowe *et al.*, 1997], but unlike GSTM4-4 and GSTM5-5 they are expressed at significant levels in selective tissues and are therefore better represented in the EST and UniGene databases. GSTM1-1 is one of the major liver isoenzymes [Strange *et al.*, 1989; Rowe *et al.*, 1997], GSTM2-2 is abundant in skeletal muscle [Laisney *et al.*, 1984; Suzuki *et al.*, 1987] and GSTM3-3 is primarily expressed in the testis and brain [Rowe *et al.*, 1997]. As discussed in Chapter 3, a large number of full-length cDNA sequences derived from diverse sources is required to enable detection of polymorphisms in low expressing genes.

A number of GST Mu class variants were also detected in various SNP databases. As was found with the sequence database-mining methods, the *GSTM5* gene was poorly represented in these SNP databases suggesting that this gene is not highly polymorphic. The one coding region variant and the few 3'-UTR and intronic variants detected (data not shown) were only found in two SNP databases: HGBASE and refseq, and by the SNPper program. Due to the low expression levels and the paucity of information available about this gene it is possible that the generation of more sequence information will reveal the existence of other *GSTM5* polymorphisms.

As with the Alpha class GSTs, problems relating to sequence identity assignations were encountered when analysing data within the public SNP databases and some UniGene clusters. *GSTM1* and *GSTM2* gene sequences were commonly confused, resulting in the false identification of a variation at residue 210 in both genes, p.S210T in *GSTM1* and p.T210S in *GSTM2*. This residue is in fact one that distinguishes *GSTM1*, which encodes p.S210, from *GSTM2*, which encodes p.T210. In addition, polymorphisms in the *GSTM4* gene verified by the CGAP were found to be *GSTM1* polymorphisms. Sequences found at the same site listed as being *GSTM1* were of an unknown identity and mapped to chromosomes 19 and X. These problems highlight the necessity to be aware of the similarities that exist between members of multi-gene families.

Overall, each SNP detection method was able to detect variants not identified by other methods. Three novel variants altering amino acids were identified and verified in the

coding regions of the *GSTM2* and *GSTM3* genes using the BLAST alignment tool and SNP Finder program (Table 4.1) and seven novel variants (excluding silent polymorphisms) were detected in all five GST Mu genes in the SNP databases (Table 4.2). Four sequence discrepancies were also detected in the EST database by the BLAST alignment tool. Of these, the *GSTM5* p.M135L and the silent *GSTM2* p.N74N variants were subsequently detected independently in other databases (Table 4.2), whilst the *GSTM3* p.W147G and silent *GSTM4* p.L55L variants had been previously noted [Comstock *et al.*, 1993; Pearson *et al.*, 1993; Ross & Board, 1993; Zhong *et al.*, 1993a]. In addition, three *GSTM4* sequence discrepancies were detected by the BLAST alignment tool: p.T4I, p.D37G and p.L63L. These were found to be sequence anomalies between the *GSTM4* sequences published by Zhong and colleagues [1993a] and Comstock and colleagues [1993]. As these have only been observed in the sequence published by Zhong and colleagues [1993a], it is likely that these represent sequencing errors.

Of the ten GST Mu class variants found, only two *GSTM3* polymorphisms were identified in the three ethnic populations tested: the p.W147G and p.V224I polymorphisms. Some variants not detected during the population screening experiments may represent rare alleles. Others may be specific to ethnic populations not screened here, however as the ethnic origin of the cDNA in the EST and UniGene databases is unknown, the relevant population cannot be identified. In contrast, frequencies calculated by the LeeLab at http://www.bioinformatics.ucla.edu/snp for the *GSTM2* p.Q129E, p.A130E and p.M134K variants and the *GSTM4* p.V212M variant indicate that these variants are relatively common (*GSTM2*: p.Q129E, C=0.74, G=0.26; p.A130E, C=0.71, A=0.29; p.M134K, T=0.75, A=0.25; *GSTM4* p.V212M, G=0.8, A=0.2). The LeeLab identified these variants in the UniGene database using a combination of various computer based analysis programs and Bayesian inference [Irizarry *et al.*, 2000]. It would appear however, that these variants were not validated in any population sets and that the allele frequencies were calculated based purely on the EST data in the UniGene database. In the absence of such validation studies, these frequencies should not be considered representative of the true situation.

The two *GSTM3* polymorphisms identified in this study have been reported elsewhere. The p.W147G polymorphism has previously been noted as a sequence alteration

unlikely to have any effect on the function of the protein [Pearson *et al.*, 1993; Ross & Board, 1993]. The existence of the *GSTM3* p.V224I polymorphism discovered using the SNP detection methods described above was independently confirmed by two groups undertaking large scale sequencing efforts, one on a subset of genes associated with neurodegenerative disease [Emahazion *et al.*, 1999] and the other interested in testing a new genome search engine [Iida *et al.*, 2001]. However, further investigations were not undertaken into characterising either of these polymorphisms. As a third polymorphism is known in the *GSTM3* gene, IVS6+22-24delAGG [Inskip *et al.*, 1995], the Australian European, Bantu African, Creole African and Southern Chinese populations were also screened for this polymorphism out of interest.

Determination of allele frequencies revealed variations in the allele distribution between the four ethnic groups, with the Southern Chinese population showing unique allele distributions. Only the Southern Chinese population encoded the extremely rare p.W147 variant, with three heterozygous individuals detected. Rare variants, those that occur at a frequency below 5%, are more likely to be restricted to specific ethnic groups [Cargill *et al.*, 1999; Halushka *et al.*, 1999; Risch, 2000] hence it is possible that the p.W147 variant is unique to the Southern Chinese population. Two of these individuals were also homozygous for the *GSTM3* p.I224 polymorphism, but the other showed heterozygosity at residue 224, leading to the conclusion that the p.W147 residue is always seen in combination with the p.I224 variant. However, as the p.W147 residue has also been seen in combination with the p.V224 residue in published *GSTM3* sequences [Campbell *et al.*, 1990; Patskovsky *et al.*, 1999c], this finding may only be true for the Southern Chinese population. Alternatively, these results may be confounded by the population size tested here. Testing larger population sets is required in order to determine whether the p.W147 variant is unique to the Southern Chinese population and to determine linkage of this polymorphism with the polymorphism at residue 224. Unlike the two African and the Australian European populations, where the p.V224 residue was observed at a frequency of approximately 66%-89%, the p.I224 variant was found at a frequency of 74% in the Chinese population. Unusually, the entire Chinese population tested here was homozygous for the known the *GSTM3*A* allele. Another study has found similar results in the Asian/Pacific Island population and although the *GSTM3*B* allele was detected, it was extremely rare (*A/*A = 98.1%, *A/*B = 1.9%, *B/*B = 0%) [Cortessis *et al.*, 2001]. Both the known *GSTM3*A* and

*GSTM3*B* alleles were seen in the Australian European and two African populations, although the *GSTM3*B* allele was much more frequent in the Bantu African samples, at 80%. The Australian European frequencies were comparable to those reported in a British study in which the polymorphism was defined (*A/*A = 70.9%, *A/*B = 25.8%, *B/*B = 3.4%) [Inskip *et al.*, 1995] and an American study (*A/*A = 58%, *A/*B = 37%, *B/*B = 5%) [Park *et al.*, 2000], as were the Creole African frequencies when compared to frequencies determined for African-American subjects (*A/*A = 37.3%, *A/*B = 44.6%, *B/*B = 18.1%) [Cortessis *et al.*, 2001]. However African-American frequencies determined in another study were different to those reported by Cortessis and colleagues [2001], and were found to be more comparable to the Bantu African frequencies typed in this study (*A/*A = 11%, *A/*B = 41%, *B/*B = 48%) [Park *et al.*, 2000]. This demonstrates the difficulty that exists when defining racial groups, for differences in prevalence rates are not only observed between similar racial groups located in different countries as shown here (Bantu and Creole Africans compared to American Africans), but are also observed in similar racial groups in the same country due to regional differences (Bantu Africans compared to Creole Africans) [Kawajiri *et al.*, 1990; Park *et al.*, 2000]. Based on the genotyping results, there appears to be no apparent association between the intronic deletion with either the p.W147G or p.V224I polymorphisms. Unfortunately, definitive associations could not be made, but these could be further clarified using allele specific amplification.

Significant changes in the catalytic ability of the GSTM3-3 protein [GSTM3A: p.G147;V224] were observed when the p.W147 (GSTM3C) and p.I224 (GSTM3B) variants were introduced singly and in combination (GSTM3D). Overall, the results in Table 4.8 demonstrate that the p.I224 residue was responsible for a significant change in enzyme activity and that this was dependent upon the amino acid encoded at residue 147 (GSTM3B, [p.G147;I224]; GSTM3D, [p.W147;I224]), whereas the substitution of the p.G147 residue for the p.W147 residue in combination with the p.V224 residue (GSTM3A, [p.G147;V224] and GSTM3C, [p.W147;V224]) only had a very small effect on the protein function. The latter substitution did not appear to dramatically alter the specific activity of GSTM3-3, only causing a 1.3- to 1.5-fold decrease in specific activity (Table 4.8). However, analysis of the kinetic parameters using CDNB as a substrate revealed an overall two-fold decrease in catalytic efficiency (K_{cat}/K_m) in GSTM3C for CDNB, brought about by a two-fold decrease in K_{cat} (Table 4.9). This,

along with the unchanged K_m value for CDNB, indicated that the protein retains its ability to bind this substrate in the presence of the p.W147 residue, but catalysis proceeds less efficiently. In contrast, the slight increase in the K_m for GSH, provided evidence for enhanced GSH binding. Combining the p.I224 residue with the p.G147 residue (GSTM3B) marginally increased the specific activity of the GSTM3A protein for the substrates CuOOH and *t*-nonenal, but caused a significant increase in activity towards the aromatic substrate CDNB (Table 4.8). This was due to a 2.9- and 8.5-fold increase in the catalytic efficiency for CDNB and GSH respectively, caused by a three-fold increase in K_{cat}. No difference in the K_m for CDNB was observed, implying that the binding affinity for this substrate had not altered however, the K_m for GSH decreased by approximately 50%, implying enhanced affinity for this substrate (Table 4.9). In contrast, when the p.I224 residue was combined with the p.W147 residue (GSTM3D), the specific activity of the GSTM3C protein with the three substrates tested was halved (Table 4.8). In this case, kinetic analyses with CDNB revealed an unaltered catalytic efficiency for CDNB, but a very slight 1.3-fold increased catalytic efficiency for GSH brought about by a combination of a small 1.2-fold decrease in K_m for GSH, and a 1.1-fold increase in K_{cat}. This implies that p.I224 in combination with p.W147 marginally increases the proteins affinity for GSH and therefore the conjugation between GSH and the electrophilic substrates would also be slightly increased. The kinetic data presented here is generally comparable to that reported for previous kinetic parameters determined for the GSTM3A protein [Rowe *et al.*, 1997; Patskovsky *et al.*, 1999c]. These studies both reported an approximate eight-fold lower K_m for GSH than found here, although K_m values for CDNB were relatively comparable (2.8 mM and 1.1 mM), as were the K_{cat} values. The catalytic efficiencies however varied between the literature and with the measurements obtained in this study, perhaps due to variations in the calculated molecular mass. It is possible that the hexyl-glutathione used to elute the GSTM3-3 proteins may not have been fully removed from the G-site during dialysis and hence competed with GSH binding, although attempts to prevent this by extensive dialysis were carried out. Differences in laboratory environments may also have contributed to these variations. Despite the inter-laboratory differences, specific activities measured here were consistent between the different batches of proteins purified, and the results are intended to show relative differences in activity.

Overall, the kinetic results are in keeping with the predictions based on 3-D structural modelling. The kinetic data show that the p.W147 residue decreases the catalytic efficiency of GSTM3-3 isoenzymes encoding the p.V224 residue (GSTM3C) for CDNB, despite previous assumptions that this polymorphism would have no functional effect on the protein [Pearson *et al.*, 1993; Ross & Board, 1993]. Predictive modelling indicates that the despite the substitution of tryptophan, a bulky residue with a non-polar ring structure side chain for glycine, a small uncharged polar residue at residue 147, relatively minor conformational changes were created in the protein structure (Figure 4.5A). Although alterations were introduced in the area neighbouring residue 147 (Figure 4.5B), its external orientation on the loop between the α5 and α6 helices, which is located away from the active sites [Raghunathan *et al.*, 1994], makes it unlikely that this residue exerts any direct effect on the protein function. However, the reduced flexibility of the environment around residue 147 that is caused by the introduction of tryptophan, and the collective differences in residue interactions in this environment, may indirectly influence the ability of the protein to easily adopt different conformations during catalysis. In addition, as seen in Figure 4.5C, conformational changes are introduced in the environment surrounding the catalytically essential p.Y11 residue, resulting in the introduction of a new H-bond, changes which could affect the ability of GSH to enter or orient itself in the G-site of the protein, and explain the slight enhancement in GSH binding affinity observed. Although kinetic parameters were not measured for CuOOH and *t*-nonenal due to time constraints, it is likely that changes in the catalytic efficiency towards these substrates would also be observed.

Introduction of the p.I224 variation into the GSTM3-3 protein also influenced the catalytic efficiency of GSTM3-3 towards CDNB. When combined with the p.G147 residue (GSTM3B), the catalytic efficiency was increased. However, considering that residue 224 is located in the three amino acid C-terminal extension unique to GSTM3-3 [Patskovsky *et al.*, 1999c], and the fact that this extension is oriented away from the protein and the active site, it would seem unlikely that this conservative substitution would have a large effect on either the function or the structure of the protein. In addition, a previous study has shown that the stability of GSTM3-3 was increased when the extension was removed, although the catalytic properties of the enzyme remained unchanged [Patskovsky *et al.*, 1999c]. Predictive modelling however, demonstrates that this substitution caused moderate conformational changes to the structure of GSTM3-3

(Figure 4.6A). It is possible that the p.I224 residue alters the stability of the protein, hence influencing the ability of substrates to access and/or fit into the active site. Notably, the conformation of the catalytically essential p.Y11 residue was considerably altered, resulting in a new H-bond interaction between p.Y11 and a neighbouring p.L17 residue (Figure 4.6C), which is believed to make up part of the catalytic site [Patskovsky *et al.*, 1999c]. This could effectively alter the dynamics of GSH binding, perhaps explaining the enhanced binding affinity observed for GSH. Although kinetic parameters were not measured for CuOOH or *t*-nonenal, and no change in specific activity was observed, it is possible that the ability of GSTM3B to catalyse these two substrates may also be affected by the presence of the p.I224 polymorphism.

Combining p.I224 with p.W147 (GSTM3D) appeared to marginally increase the binding ability of the GSTM3C protein [p.W147;V224] for GSH. Predictive modelling demonstrated that the combined effect of these variants were not responsible for any major structural differences to GSTM3-3. p.W147 showed similar conformational differences when combined with p.I224, as when seen in combination with p.V224 (Figure 4.5B). In addition, the conformational alterations observed around p.Y11 were similar to that seen in GSTM3C [p.W147;V224] (Figure 4.5C). As discussed above, these changes could affect the ability of GSH to enter or orient itself into the G-site of the protein. In contrast, the orientation of p.I224 in combination with p.W147 was very different to that adopted in combination with p.G147, resulting in different local conformational alterations. Again, as discussed above, this could possibly affect the stability of the protein. Although kinetic parameters were not measured for CuOOH and *t*-nonenal, it is likely that the decreased specific activities observed are also due to similar effects.

GSTM3-3 possesses the highest peroxidase activity of the GST Mu class family towards organic hydroperoxides and, unlike the Alpha class GSTs, is also able to reduce H_2O_2 [Campbell *et al.*, 1990; Comstock *et al.*, 1994]. Since a two-fold decrease in specific activity towards CuOOH was observed in GSTM3-3 containing both the p.W147 and p.I224 polymorphisms (GSTM3D), it is possible that an increased susceptibility to disease related to oxidative stress might be observed in individuals carrying these alleles. On the basis of the kinetic studies reported here, it can also be speculated that GSTM3-3 proteins encoding only one of these alleles may also alter the

peroxidase activity of GSTM3-3, with the p.W147 residue decreasing the catalytic efficiency and the p.I224 residue increasing catalytic efficiency. However, due to the lack of kinetic data confirming these speculations, the individual contribution of these polymorphisms to the peroxidase activity of GSTM3-3 will not be considered further. Disease or damage related to altered GSTM3-3 activity is most likely to be manifested in the testes and brain, both of which possess a blood barrier and high expression levels of GSTM3-3 [Campbell *et al.*, 1990; Rowe *et al.*, 1997]. GSTs expressed at these barriers are believed to provide the first line of defence against endogenous compounds and xenobiotics [Abramovitz *et al.*, 1988; Cammer *et al.*, 1989; Beuckmann *et al.*, 2000]. As brain cells are not able to regenerate and germ cell generation requiring high fidelity DNA replication is constantly occurring in the testes, polymorphisms affecting the function of the GSTs that protect these organs may alter susceptibility to irreparable cellular damage.

The effects of the decreased peroxidase activity exhibited by the variant GSTM3D protein may also affect the reduction and isomerisation of the prostaglandin PGH_2 to PGE_2 in the brain, a reaction catalysed by GSTM3-3 [Beuckmann *et al.*, 2000]. PGE_2 is one of the two major arachidonic acid metabolites in mammalian neurons and glial cells and plays several neurophysiological roles in the central and peripheral nervous systems, including neuroendocrine function (sedation, catatonia, hypertension), body temperature regulation including mediation of fever [Coceani & Akarsu, 1998] and hyperthermia [Onoe *et al.*, 1992], nociception and allodynia [Minami *et al.*, 1994]. It is also involved in the regulation of sleep-wakefulness, where in combination with the sleep inducer PGD_2, it promotes wakefulness [Onoe *et al.*, 1992]. The high expression levels of GSTM3-3 in the brain imply that this isoenzyme may be of importance in the production of PGE_2 in the brain [Beuckmann *et al.*, 2000]. It can therefore be postulated that polymorphisms in GSTM3-3 that affect peroxidase activity may directly interfere with PGE_2 synthesis and indirectly influence the various roles of PGE_2.

The Mu class GSTs have a protective role in detoxification of the nitrosoureas, a group of alkylating antineoplastic drugs, and over-expression of the Mu class GSTs is suggested to be involved in the development of resistance to the nitrosoureas in tumour cells [Tew, 1994]. For example, increased expression of GSTM3-3, along with GSTM1-1, has been measured in human lung cancer lines resistant to doxirubicin and

cisplatin [Hao *et al.*, 1994]. Of particular interest however, is the finding that GSTM3-3 over-expression inactivates the chloroethyl nitrosourea BCNU [Berhane *et al.*, 1993; Egyházi *et al.*, 1997], the drug of choice for treating brain tumours. The implications of polymorphisms within GSTM3-3 are vast, considering it is found expressed at high levels in the brain, and these polymorphisms could possibly alter resistance to this hydrophobic drug.

When considering the effects the GSTM3-3 polymorphisms may have on protein function, it is important to take interindividual variation of protein expression into account, as discussed in Chapter 3. Various studies have demonstrated distinct interindividual variation of GSTM3-3 expression within tissues [Anttila *et al.*, 1995; Nakajima *et al.*, 1995] and it is thought that this may be regulated by the intronic *GSTM3* polymorphism encoded by the *GSTM3*B* allele due to the presence of the YY1 transcription factor in *GSTM3*B* [Hariharan *et al.*, 1991; Shi *et al.*, 1991; Inskip *et al.*, 1995]. Another possibly more important consideration to take into account however, is the linkage disequilibrium that exists between the *GSTM3*B* and *GSTM1*A* alleles [Inskip *et al.*, 1995], an association believed to increase the expression, and thus the detoxification abilities, of GSTM3-3 [Yengi *et al.*, 1996]. Through linkage with *GSTM1*A*, *GSTM3*B* is believed to confer protective effects on disease normally associated with the *GSTM1*0* allele such as skin cancers [Yengi *et al.*, 1996] and various smoking related diseases such as lung [To-Figueras *et al.*, 2000] and oral cancers [Jahnke *et al.*, 1996; Matthias *et al.*, 1998]. Oral cancer association studies have, however, been conflicting. No associations have been found between the *GSTM1* and *GSTM3* alleles with oral or pharyngeal cancers [Jourenkova-Mironova *et al.*, 1999a]. Ethnic differences have been observed in the incidence of laryngeal cancer [Worrall *et al.*, 1998; Park *et al.*, 2000] and an increased risk of laryngeal cancer has been reported due to the combined effects of the *GSTM1*0* and *GSTM3*A*B* or *GSTM3*B*B* alleles [Jourenkova-Mironova *et al.*, 1999b]. It has been suggested that the negative and positive regulatory properties of YY1 may be responsible for this conflict [Jouvenkova-Mironova *et al.*, 1999b]. It may be possible that the two *GSTM3* polymorphisms detected in this study are also strongly associated with the expression of specific *GSTM1* alleles, the *GSTM3* intronic deletion or with other genes and thereby act as risk modifiers to various diseases. It would be of interest to determine whether the two *GSTM3* polymorphisms detected in this study can be associated with the

GSTM1 polymorphisms or the *GSTM3* intronic polymorphism and if they are, to determine whether these associations exert any functional effects or affect expression levels.

A final point to take into consideration when determining the effects the *GSTM3* polymorphisms may have on protein function is the large variation in the distribution of allele frequencies among the ethnic groups studied here. The Chinese population presented a completely different *GSTM3* polymorphism profile to the Australian European and two African populations, and whilst this profile may be beneficial or contribute to a disease phenotype in the Southern Chinese population, the same profile in the Bantu African, Creole African or Australian European populations may be associated with different phenotypes. For example, a significant increase in susceptibility to head and neck cancer associated with the *GSTM1**0 allele has been observed in the Japanese population, but not in Caucasians [Kihara *et al.*, 1993; Park *et al.*, 1997]. Also, as discussed earlier, the distribution of allele frequencies is variable both between races and also within racial groups. This regional and international variation is influenced by differences in genetic, environmental or dietary interactions [Kawajiri *et al.*, 1990; Garte, 1998; Park *et al.*, 2000; Risch, 2000].

Using a combination of database-mining methods, a moderate level of genetic polymorphism was identified in the human Mu class GST genes. Although each method yielded results, this chapter has shown that identification of polymorphism should not be limited to one method or technique, especially since some genes are not as well represented as others in the various databases. Two polymorphisms exhibiting functional and structural differences were verified in three ethnic populations, the GSTM3-3 p.W147G and p.V224I polymorphisms. Further characterisation is required to determine the physiological effects of each.

CHAPTER 5

THE OMEGA CLASS GSTS

5.1 INTRODUCTION

The two Omega class GSTs, *GSTO1* and *GSTO2*, represent the most recently discovered family of GST. Unlike many of the GSTs, which were discovered using traditional protein purification and characterisation methods, the Omega class GSTs were discovered through analysis of the EST database and the GenBank genomic sequence database [Board *et al.*, 2000; Morel *et al.*, 2001; Whitbread *et al.*, In Press]. Difficulties with purifying the GSTO2-2 isoenzyme and a lack of GSTO1-1 activity with the typical range of GST substrates has made it difficult to characterise the role these GSTs may play in detoxification. Despite this, GSTO1-1 has been ascribed novel roles not seen before in any GST isoenzyme including weak thiol transferase and dehydroascorbate reductase activities, monomethylarsonic acid reductase activity and regulation of ryanodine receptor activity, which together with the broad expression profile suggests a fundamental role in cellular metabolism [Board *et al.*, 2000].

When initially described, it was found that GSTO1-1 possessed weak thiol transferase and dehydroascorbate reductase activities [Board *et al.*, 2000] normally associated with the glutaredoxin enzymes, which allows them to provide protection against oxidative damage [Holmgren, 1989; Holmgren, 2000]. Using their thiol-transferase activity, the glutaredoxins, and possibly GSTO1-1 [Board *et al.*, 2000], are able to reduce *S*-thiolated proteins and thus restore enzyme activity [Flamigni *et al.*, 1989; Seres *et al.*, 1996]. *S*-thiolation of proteins causes conformational changes and hence alters the activity of the thiolated protein, a process resulting from oxidative stress in order to provide protection against irreversible oxidative damage [Kono *et al.*, 1990; Coan *et al.*, 1992; Del Corso *et al.*, 1994; Seres *et al.*, 1996; Hanson *et al.*, 1999]. Using this dehydroascorbate reductase activity, the glutaredoxins, and again possibly GSTO1-1, are able to provide enough ascorbate to protect neutrophils from the oxidants they release when activated in order to protect against bacterial infections [Halliwell *et al.*, 1987; Washko *et al.*, 1993; Park & Levine, 1996].

GSTO1-1 has also been implicated in the four-step methylation of the carcinogenic compound inorganic arsenic (Figure 5.1). Arsenic has been commonly used in pesticides, rodenticides, herbicides and in industry, and is generated during mining and

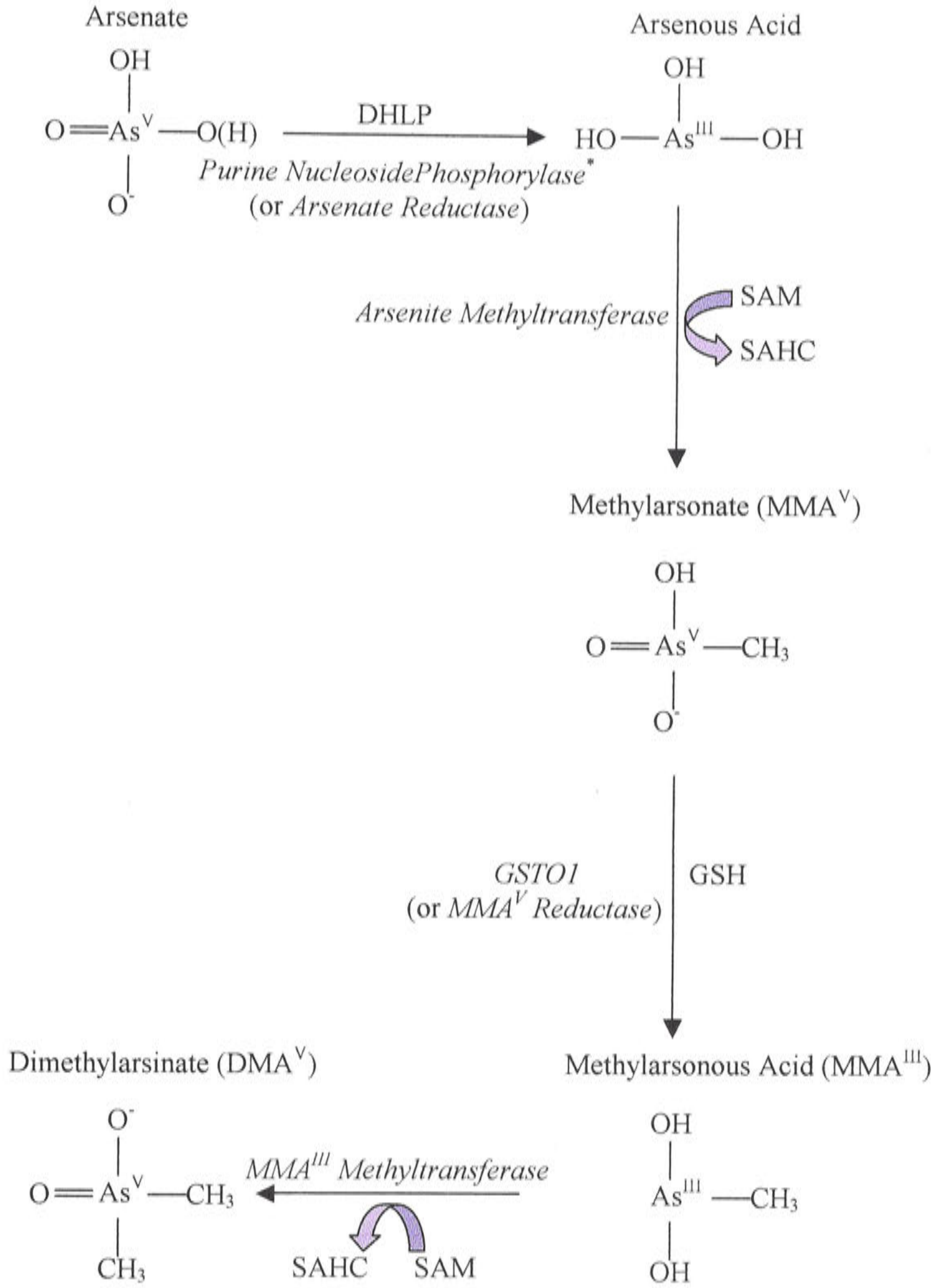


Figure 5.1 – The inorganic arsenic biotransformation pathway (adapted from [Zakharyan *et al.*, 2001]). SAM is *S*-adenosyl-L-methionine; SAHC is *S*-adenosyl-L-homocysteine. (*) Purine Nucleoside Phosphorylase has recently been identified as Arsenate reductase. This reaction requires (and prefers) DHLP (dihydrolipoic acid) to proceed, not GSH as previously believed [Radabaugh *et al.*, 2002].

smelting activities and burning of fossil fuels [Aposhian, 1997; Smedley & Kinniburgh, 2001]. Inorganic arsenic contaminates the drinking water of many countries including Bangladesh and some Central American, North American, Asian and Eastern European countries in excess of the maximum permissible concentrations stipulated by the World Health Organisation (WHO) [Smedley & Kinniburgh, 2001; Yamamura *et al.*, 2001] and has been associated with various cancers, in particular skin cancer [Bates *et al.*, 1992], vascular disorders such as Blackfoot disease [Tseng, 1977], hyperkeratosis, skin pigmentation changes, neurotoxicity and hepatoxicity [Abernathy & Morgan, 2001]. Paradoxically, arsenical compounds have been used medicinally for thousands of years [Klaasen, 1996] and arsenic trioxide is currently used as an effective chemotherapeutic agent against many cancers, particularly AML [Soignet *et al.*, 1998], although concerns have recently arisen over the use of this compound due to some unexpected deaths of acute promyelocytic leukemia patients undergoing arsenic trioxide therapy [Westervelt *et al.*, 2001]. Recently, GSTO1-1 was found to be 100% identical to human liver MMA^{V} reductase, which is responsible for catalysing the GSH-dependent reduction of MMA^{V} to MMA^{III}, the rate limiting penultimate GSH-dependent step in this pathway [Zakharyan & Aposhian, 1999; Zakharyan *et al.*, 2001]. The product, MMA^{III}, has been found to be more toxic than the parent compound arsenite in mammalian tissues, suggesting that this may be responsible for the pathological effects usually attributed to arsenic [Petrick *et al.*, 2000; Sampayo-Reyes *et al.*, 2000]. Tolerance and variability in the metabolism of inorganic arsenic have been reported [Roscoe, 1862; Vahter *et al.*, 1995; Aposhian, 1997]. It has been suggested that polymorphisms in the enzymes involved in the arsenic biomethylation pathway, especially in the rate-limiting MMA^{V} reductase, may be responsible for this observed variability [Zakharyan & Aposhian, 1999]. It has also been suggested that variability in the toxicity of arsenic trioxide used for treatment of AML may have a genetic basis [Westervelt *et al.*, 2001].

Unlike any of the other characterised GST isoenzymes, a recent study has shown that GSTO1-1 is able to non-catalytically modulate the activity of ryanodine receptor (RyR) calcium release channels found in the endoplasmic reticulum [Dulhunty *et al.*, 2001]. More specifically, GSTO1-1 was able to potentiate the activity of the skeletal muscle associated ryanodine receptor 1 (RyR1) and inhibit the activity of the cardiac muscle associated ryanodine receptor 2 (RyR2) [Dulhunty *et al.*, 2001]. RyRs are responsible for regulating the release of Ca^{2+} from intracellular stores. Activation of RyRs during

oxidative stress is responsible for an increase in cytoplasmic Ca^{2+} levels, an event that can ultimately trigger apoptosis in both normal and tumour cells [Berridge *et al.*, 1998; Mariot *et al.*, 2000; Pan *et al.*, 2000]. It has been suggested that through modulating activity of RyR1 and RyR2, GSTO1-1 may be involved in conferring protection against apoptosis and oxidative stress [Dulhunty *et al.*, 2001].

This chapter describes the use of various database mining techniques to facilitate the discovery of novel Omega class polymorphisms. The distribution of the confirmed polymorphisms was determined in three ethnic populations and the recombinant proteins were characterised enzymatically to determine whether the different sequence variations had an effect on the function of the enzyme.

5.2 Materials and Methods

Research into the Omega class GSTs was carried out in collaboration with Astrid Whitbread, a PhD student in the Molecular Genetics Group. Specifically, all recombinant protein work and enzymatic characterisation was carried out by Astrid Whitbread. Database analysis, PCR/RFLP and structural analysis of the different GSTO1-1 variants was carried out by the author. Database analysis and standard PCR methods used in this chapter are described in Chapter 2. Techniques unique to this chapter are outlined below.

5.2.1 Database Screening

Potential polymorphisms in the two Omega class GSTs were identified and verified using the database mining methods described in §2.3, and are listed in Tables 5.1 and 5.2.

5.2.2 PCR/RFLP Analysis

Potentially polymorphic exons of the *GSTO1* and *GSTO2* genes were specifically amplified from genomic DNA using PCR as described in §2.5.6.2, in order to perform small-scale population studies. The oligonucleotide sequences and the specific conditions for each PCR are listed in Table 5.3. The presence of variation in the PCR products was determined by RFLP analysis (§2.5.4.2) using the appropriate restriction enzymes (Table 5.4). All variants detected altered restriction enzyme sites. Up to 100 samples from the Australian European, Bantu African, Creole African and Southern Chinese population groups were amplified by PCR to allow allele and haplotype frequency determination for polymorphisms confirmed by the small-scale population screens.

Table 5.1 – GST Omega variants detected in the EST database by the BLAST alignment tool and in the UniGene database by the SNP Finder program.

Gene	Nucleotide	Residue	Exon	Program	ESTs with alteration	cDNA libraries represented	Sequence confirmed[a,b]	Confirmed in population studies
GSTO1	c.98C>G	p.P33R	2	BLAST	2	2	No-AI057541	N.S.
	c.169A>G	p.K57E	3	BLAST	2	2	No-AA889961	N.S.
	c.173A>T	p.N58I	3	BLAST	2	2	No-AI084788[c]	No
	c.229A>G	p.N77D	3	BLAST	2	1	No-AI933306	N.S.
	c.252C>G	p.Y84X	3	BLAST	2	2	No-AW150107	N.S.
	c.277C>G	p.L93V	3	BLAST	2	2	No-AI222733	N.S.
	c.297G>A	p.G99G	3	BLAST	2	2	No-AW150107	N.S.
	c.311C>G	p.P104R	3	BLAST	2	2	No-AW150107	N.S.
	c.388T>C	p.F130L	4	SNP Finder	3	3	Yes-0.20	No
	c.419C>A	p.A140D	4	BLAST	20	16	Yes-BE385940	Yes
	c.423C>G	p.G141G	4	SNP Finder	2	2	No-0.20	N.S.
	c.445G>T	p.E149X	4	BLAST	2	2	Yes-BE389540	No
	c.495T>A	p.F165L	5	SNP Finder	2	1	Yes-0.99	No
	c.550InsG	Frameshift	5	BLAST	2	2	No-BE622343	N.S.
	c.585C>A	p.H195Q	6	BLAST	2	2	No-BE560141	N.S.
	c.587C>A	p.T196N	6	BLAST	1	1	No-BE791208	N.S.
	c.587C>T	p.T196I	6	BLAST	2	2	No-BE728901	N.S.
	c.406G>A	p.W202X	6	SNP Finder	1	1	Yes-0.96	No
	c.614C>G	p.A205G	6	BLAST	2	2	No-BE251897	N.S.
	c.636C>T	p.V212V	6	SNP Finder	11	11	Yes-0.61	N.S.
	c.650C>A	p.T217N	6	SNP Finder	1	1	Yes-0.99	No
	c.669A>G	p.Q223Q	6	BLAST	2	1	No-BE560141	N.S.
	c.699C>G	p.S233R	6	BLAST	3	3	No-BE796746	N.S.
GSTO2[d]	c.207T>G	p.F69L		BLAST	2	2	No-BE304990	N.S.
	c.424A>G	p.N142D	4	BLAST	5	5	Yes-AW301074	Yes

[a] The accession number of the I.M.A.G.E. Consortium (LLNL) cDNA clones sequenced to confirm the presence of potential polymorphisms identified using the BLAST alignment tool is provided

[b] Polymorphisms detected using the SNP Finder program are allocated a likelihood score. As the number approaches 1.0, there is a greater likelihood that this polymorphism will be real. Sequences were confirmed by viewing the sequence trace data made available at the site.

[c] The incorrect I.M.A.G.E. Consortium (LLNL) cDNA clone was received. Small-scale population screening was performed instead to validate this sequence variation.

[d] *GSTO2* was not represented in the UniGene clusters, hence could not be analysed using the SNP Finder program.

N.S.: Not Studied

Table 5.2 – *GSTO1* polymorphisms detected in the SNP databases.

Nucleotide	Residue	Exon	Database	Sequence confirmed	Confirmed in population studies
c.169A>G	p.K57E[a]	3	refseq (0.96)[b]	Yes	No
c.361T>C	p.S121P	3	dbSNP (0.01)[b]	Yes	No
c.406A>T	p.K136X	4	dbSNP (0.11)[b]	Yes	No
c.419C>A	p.A140D[a]	4	dbSNP, refseq (0.98)[b], HGBASE	Yes	Yes
c.423C>G	p.G141G[a]	4	dbSNP (0.02)[b]	No	N.S.
c.432A>G	p.E144E	4	dbSNP, refseq, HGBASE	Unable[c]	N.S.
c.445G>T	p.E149X[a]	4	refseq (0.96)[b]	Yes	No
c.462G>T	p.E154D	4	dbSNP (0.02)[b]	Yes	No
c.495T>A	p.F165L[a]	5	CGAPC (0.99), refseq (0.99)[b]	Yes	No
c.606G>A	p.W202X[a]	6	refseq (0.96)[b]	Yes	No
c.636C>T	p.V212V[a]	6	dbSNP (0.61)[b]	Yes	N.S.
c.636C>A	p.V212V	6	dbSNP (0.61)[b]	Yes	N.S.
c.650C>A	p.T217N[a]	6	CGAPC(0.99), dbSNP (0.99)[b], refseq (0.99)[b], HGBASE	Yes	No
c.723C>T	p.L241L	6	dbSNP, refseq	Unable[c]	N.S.

[a] These polymorphisms were also detected by the BLAST alignment tool and the SNP Finder program.

[b] If the SNP was submitted by the GAI to different databases, links to the SNP Finder program were provided, allowing access to available sequence trace data.

[c] Some sequences were unable to be confirmed directly through the databases, therefore small-scale population screens were used to confirm these polymorphisms.

Table 5.3 – Oligonucleotides used for *GSTO1* and *GSTO2* exon amplification.

Primer	Exon	Primer sequence 5' → 3'	Annealing Conditions	Product Length
GSTO1				
O1Ex3F	3	AGCTAAGTGGATGGCAAAGCC	60°C, 18 sec	394 bp
O1Ex3R		TCACTGAACCATAAATGGCAGC		
O1Ex4F	4	TCTAGGTGCCATCCTTG	56°C, 18 sec	127 bp
O1Ex4Rev[a]		TGATAGCTAGGAGAAATAATTACCTC		
O1Ex4Rev2[b]		TGATAGCTAGGAGAAATAA	48°C, 18 sec	127 bp
O1Ex4A2	4	CGATACAGTTAGCCATAAAC	56°C, 20 sec	303 bp
O1Ex4R2		AACAGCAGTCTGCTGTAAAC		
O1Ex5F	5	GTGAGTCTTACTACATGCACAAG	63°C, 20 sec	207 bp
O1Ex5R		CCACCTGTCATCCTAAATTATGC		
O1Ex6F	6	TCATCCTAGTTGACCTAGCTCAC	57°C, 23 sec	334 bp
O1Ex6R		CTGTTGATGTCAAATCAGTCAGAG		
GSTO2				
O2Ex4F	4	GATTAGGTCCCACATTTG	50°C, 20 sec	112 bp
O2Ex4R		GTACCTCTTCCAGGTTG		

[a] This primer was used in conjunction with O1Ex4F using the annealing conditions 56°C, 18 sec.

[b] This primer was used in conjunction with O1Ex4F using the annealing conditions 48°C, 18 sec.

Table 5.4 – Restriction Endonuclease Sites used to detect polymorphisms in *GSTO1* and *GSTO2*.

Variation	Exon	Endonuclease Site	Fragment Size (bp)		
			Common Allele	Heterozygous	Variant Allele
GSTO1					
p.N58I	3	*Mse*I	221, 136, 37	221, 136, 118, 37, 18	221, 118, 37, 18
p.S121P	3	*Bsu*36I	394	394, 310, 84	310, 84
p.F130L	4	*Hind*III	104, 23	127, 104, 23	127
p.K136X	4	*Mse*I	127	127, 83, 44	83, 44
p.A140D	4	*Cac*8I	68, 59	127, 68, 59	127
p.E149X	4	*Apo*I	72, 43, 12	72, 55, 43, 12	72, 55
p.E154D	4	*Bse*RI	127	127, 114, 13	114, 13
p.F165L	5	*Dde*I	207	207, 125, 82	125, 8233
p.W202X	6	*Fok*I	238, 96	334, 238, 96	334
p.T217N	6	*Spe*I	207, 127	334, 207, 127	334
GSTO2					
p.N142D	4	*Sau*3AI	112	112, 63, 49	63, 49

5.2.3 Identification of a Deletion Polymorphism

Electrophoresis of *GSTO1* Exon 4 PCR products on polyacrylamide gels revealed the presence of three different patterns: (1) the expected 127 bp wild-type pattern, (2) a 124 bp fragment and (3) a four-band pattern, with fragments of 124 and127 bp, and a doublet of approximately 130 and 131 bp (Figure 5.2). A PCR sample with the four-band pattern was cloned into the pGEM®-T vector, following the manufacturers protocol (Promega). Clones with either the 127 bp or 124 bp insert were identified by PCR amplification as described in §2.5.6.2 using the PCR primer pair O1Ex4F and O1Ex4Rev2 (Table 5.3), followed by electrophoresis through 7% acrylamide gels (§2.5.2.2). Clones containing either the 127 bp or 124 bp insert were sequenced using the ThermoSequenase cycle sequencing kit as described in (§2.5.5.1).

5.2.4 Recombinant Protein Analysis

5.2.4.1 GSTO1-1

5.2.4.1.1 Site-Directed Mutagenesis

The GSTO1-1 protein, identified as the GSTO1*A allelic isoform (Table 5.5), was expressed in *E. coli* in the plasmid pQE30 (Qiagen). Using the QuikChange Site-Directed Mutagenesis Kit (§2.6.3), the GSTO1*B and GSTO1*C allelic isoforms were created from GSTO1*A using the primer pair O1A140DA and O1A140DB, and E155null1hO1 and E155null2hO1 respectively (Table 5.6). All mutations were confirmed by cycle sequencing (§2.5.5.1).

5.2.4.1.2 Protein Expression and Purification

Plasmids encoding the GSTO1*A, GSTO1*B and GSTO1*C allelic isoforms were transfected into *E.coli* strain M15(pRep4) (§2.6.2). To express these, 10 mL of an overnight culture was diluted 1:40 into LB supplemented with 100 µg/mL ampicillin

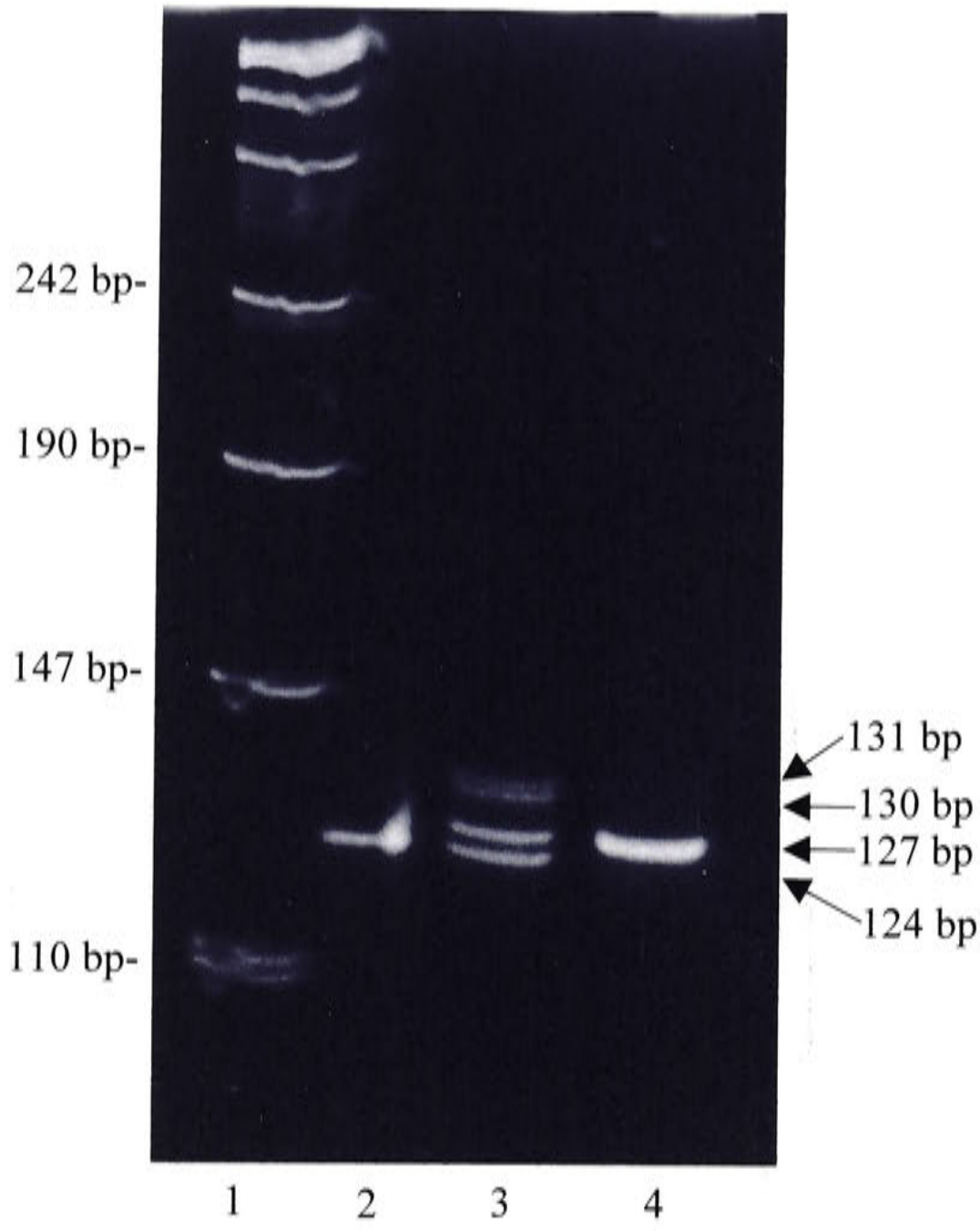


Figure 5.2 - *GSTO1* exon 4 undigested PCR products showing the presence (*GSTO1*B*), or absence (*GSTO1*A*) of the three base pair deletion spanning the exon4/intron 4 boundary. Lane 1 – pUC/*Hpa*II marker; Lane 2 – the 124 bp *GSTO1*B* homozygote; Lane 3 – the *GSTO1*A*B* heterozygote; Lane 4 – the 127 bp *GSTO1*A* homozygote.

Table 5.5 – Nucleotide and amino acid variations in *GSTO1*.

Haplotype	**Nucleotide**	
	c.419	c.464-IVS+1
GSTO1*A	C	aag
GSTO1*B	C	Deleted
GSTO1*C	A	aag
GSTO1*D	A	Deleted
	Residue	
	p.140	p.155
GSTO1*A	Ala	Glu
GSTO1*B	Ala	Deleted
GSTO1*C	Asp	Glu
GSTO1*D	Asp	Deleted

Table 5.6 – Mutagenesis primers used to create GSTO1-1 recombinant protein variants.

GSTO1-1 Haplotype	Primer Name	Mutagenesis Primer Sequences 5' to 3'
GSTO1B	GSTO1A140D1	GCCAAAATAAAGAAGACTATGATGGCCTAAAAGAAGAATTTCG
	GSTO1A140D2	CGAAATTCTTCTTTTAGGCCATCATAGTCTTCTTTATTTTTGGC
GSTO1C	E155null1hO1	GAATTTACCAAGCTAGAGGTTCTGACTAATAAGAAAGACGACC
	E155null2hO1	GGTCGTCTTCTTATTAGTCAGAACCTCTAGCTTGGTAAATTC

and 20 μg/mL kanamycin. The 400 mL cultures were grown at 37°C to an OD_{600} of approximately 0.8 and protein expression was induced with 0.1 mM IPTG for five hours. The cells were harvested by centrifugation at 8.5 K for 10 minutes at 4°C (GSA-1500 rotor) and resuspended in 20 mL resuspension buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 6). Cells were lysed by incubation with 200 μg/mL lysozyme on ice for one hour followed by sonication (3 x 30 second bursts). Cellular debris was removed by centrifugation at10 K for 10 minutes at 4°C (SS-34 rotor).

The proteins were all purified at 4°C by affinity chromatography. Cleared lysates were incubated for one hour on a rotary mixer with nickel agarose. The agarose was collected by centrifugation at 1000 rpm for 5 minutes and washed with 50 mL of 50 mM NaH_2PO_4, 300 mM NaCl, pH 6. The agarose was washed with a further 500 mL of 50 mM NaH_2PO_4, 300 mM NaCl, pH 6 on a scintered funnel, then transferred to a column and washed with a further 500 mL of 50 mM NaH_2PO_4, 300 mM NaCl, pH 6 followed by 20 mL of 20 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.5. The proteins were eluted from the nickel agarose with 500 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.5. Fractions containing recombinant GSTO1-1 were identified by 12.5% SDS-PAGE (§2.5.2.3). The strongest fractions were pooled and dialysed in 20 mM Tris, 60 mM NaCl pH 8. Protein concentrations were determined using the Lowry method (§2.5.2.3).

5.2.4.2 GSTO2-2

5.2.4.2.1 Amplification and Isolation of the *GSTO2* cDNA

GSTO2 cDNA, later found to encode the GSTO2*A allelic isoform (p.N142), was isolated by PCR amplification of the I.M.A.G.E. Consortium (LLNL) cDNA Clone BE908662, which encodes *GSTO2*, using the forward primer hGSTO2EXF 5'-CACGGATCCTCTGGGGATGCGACCAG-3' and the reverse primer hGSTO2EXR 5'-CTGAAGCTTTCAGCACAGCCCAAAGTC-3', which have a *Bam*HI and *Hind*III restriction site incorporated respectively. A 25 μL reaction volume contained 1x ABI reaction buffer IV, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 10 pmol of the forward and reverse

primers and 0.1 unit of *Taq* DNA polymerase. A negative control sample containing all components of the reaction except the template was included in each PCR to ensure the absence of contamination. PCR conditions involved an initial denaturation step at 95°C for 2 minutes followed by 30 cycles of a 95°C denaturation step (30 sec), 54°C annealing step (30 sec), and 72°C extension step (1 min), with a final 3 minute extension at 72°C.

cDNA was also extracted from a human testis cDNA library λgt11 as described in §2.5.1.1 and *GSTO2* cDNA was amplified by PCR using the primer pair hGSTO2EXF and hGSTO2EXR described above. This was later found to encode the GSTO2*B allelic isoform (p.D142). A 25 μL reaction volume contained 1x ABI reaction buffer IV, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 5 pmol of the forward and reverse primers, and 0.5 unit of *Taq* DNA polymerase. A negative control sample containing all components of the reaction except the template was included in each PCR to ensure the absence of contamination. PCR conditions involved an initial denaturation step at 95°C for 2 minutes followed by 30 cycles of a 95°C denaturation step (30 sec), 54°C annealing step (30 sec) and 72°C extension step (1 min), with a final 3 minute extension at 72°C. Agarose gel electrophoresis was used to confirm the absence of spurious bands and contamination and the presence of PCR product (3 μL).

5.2.4.2.2 SUBCLONING

The 747 bp *GSTO2*A* PCR product was digested with the restriction enzymes *Bam*HI and *Hind*III and cloned into the *Bam*HI and *Hind*III sites of the pQE30 expression vector, as described in §2.5.4.1 and §2.6.1, downstream and in-frame of the six-histidine coding sequence. The 747 bp *GSTO2*B* PCR product was similarly treated after initially being subcloned into the pGEM®-T (Promega) PCR cloning vector. Both constructs were confirmed as the GSTO2A and GSTO2B allelic isoforms by cycle and automated sequencing as described in §2.5.5.1 and §2.5.5.2.

5.2.5 GSTO1-1 Enzyme Characterisation

5.2.5.1 Structural Analysis

The GSTO1*A [p.A140;E155], GSTO1*B [p.A140;E155del] and GSTO1*C [p.D140;E155] allelic variants were built into the 3-D model of the GSTO1-1 protein (PDB file 1EEM) as described in §2.6.5.

5.2.5.2 Enzymatic Analysis

Activity for the GSTO1*A, GSTO1*B and GSTO1*C allelic isoforms was measured with the substrate CDNB and thioltransferase activity with the substrate HEDS as described in §2.6.4. Three preparations of each isoform were prepared.

5.3 RESULTS

5.3.1 DETECTION OF OMEGA CLASS POLYMORPHISMS

5.3.1.1 DETECTION USING DATABASE ANALYSIS PROGRAMS

Polymorphisms in Omega class cDNA sequences located in the human EST and UniGene databases were detected using the BLAST alignment tool and SNP Finder program respectively. The EST database contained 100 *GSTO1* and 46 *GSTO2* cDNAs. The UniGene clusters contained 179 *GSTO1* cDNA and mRNA sequences. A cluster for the *GSTO2* gene was not available at UniGene, hence this could not be analysed using the SNP Finder program. From these sequence databases, 21 variants that altered amino acid residues were detected in both Omega genes (Table 5.1). Of these, seven were confirmed upon resequencing a representative I.M.A.G.E. Consortium (LLNL) cDNA clone or viewing existing electropherograms.

5.3.1.2 DETECTION USING SNP DATABASES

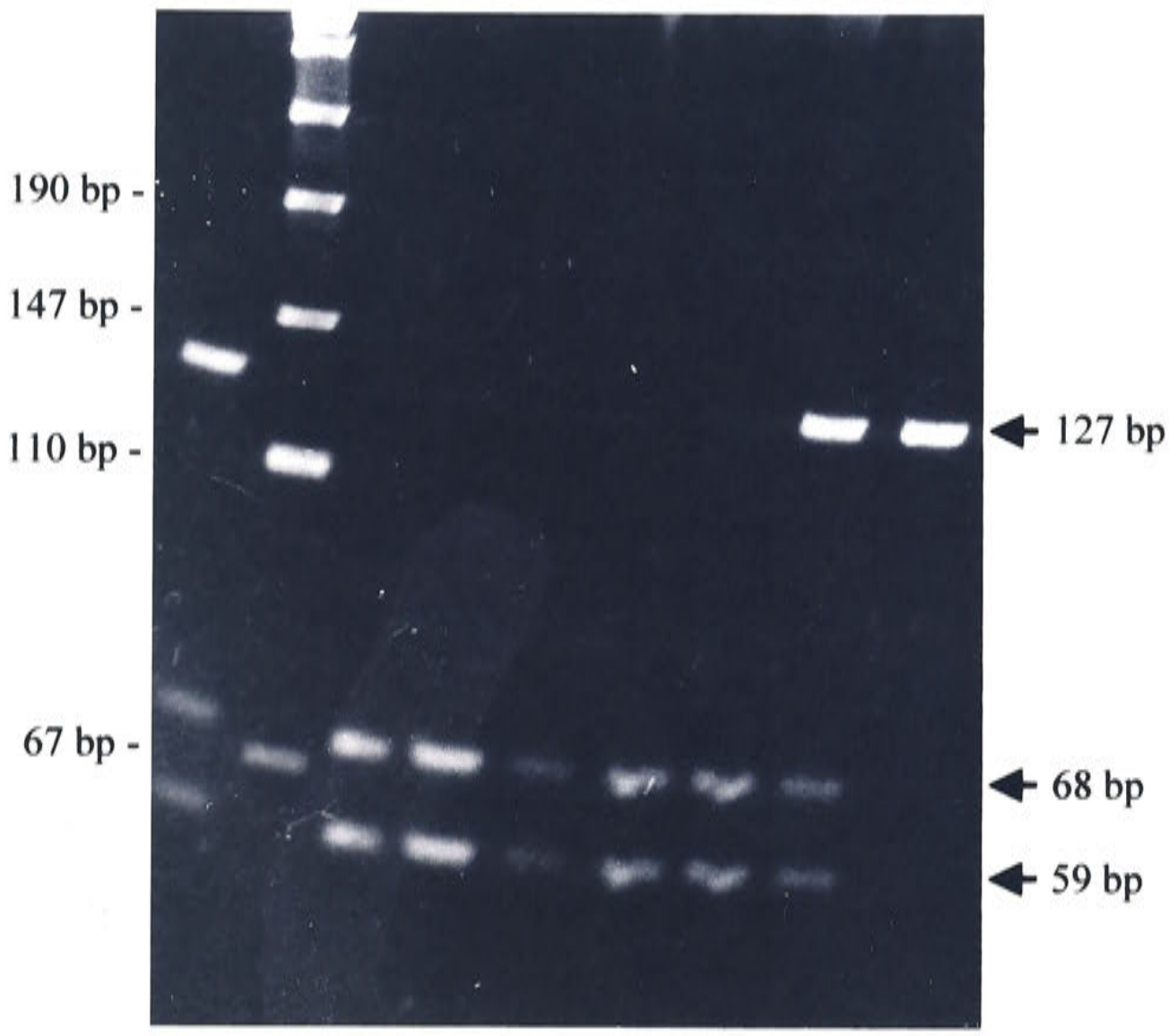
The dbSNP, HGBASE, GeneSNPs, EGPSNPs and CGAP lists were examined for the presence of *GSTO1* and *GSTO2* sequences. The SNPper program was also utilised. Many *GSTO1* submissions were found and when limited to SNPs that alter amino acid residues, three novel variants were identified (Table 5.2). Although some *GSTO1* variants could be confirmed through analysis of electropherograms, small-scale population analysis was undertaken for confirmation of all sequence variations identified. Eight variants that had been detected using the BLAST alignment tool and SNP Finder programs were also detected in these databases. *GSTO2* was not found in any of the databases analysed.

5.3.1.3 Determination and Distribution of Allele Frequencies

25 samples from three ethnic groups were tested for the presence of the ten variants. Determination of the nucleotide present at variant sites was achieved by amplification of the relevant exons followed by digestion with the appropriate restriction enzyme (Table 5.4). Of these, only two polymorphisms: *GSTO1* p.A140D (Figure 5.3) and *GSTO2* p.N142D (Figure 5.4), were found in the sample populations. Allele frequencies for these two polymorphisms were subsequently determined in up to 100 samples for the Australian European, Bantu African, Creole African and Southern Chinese populations (Table 5.7). For the *GSTO1* polymorphism, the p.D140 allele was relatively rare in the two African and Southern Chinese populations, being found at frequencies between 10-18%. Homozygotes for this allele were absent in the Bantu African population. In contrast, the p.D140 allele occurred more frequently in the Australian European population, at 33.5%. Chi-square analysis provided support for genotypic differences between the four populations ($\chi^2_6 = 31.58$, $p<0.05$). Notably, evidence was provided for a genotypic difference between the Australian European population with the remaining three groups (Australian European/Bantu African: $\chi^2_2 = 24.49$, $p<0.05$; Australian European/Creole African $\chi^2_2 = 10.53$, $p<0.05$; Australian European/Southern Chinese: $\chi^2_2 = 14.59$, $p<0.05$). In *GSTO2*, the p.D142 allele was found to be quite common, and occurred most frequently in the two African populations, especially in the Bantu African population where it appears that this is the common allele (p.D142=83%). Evidence for significant genotypic differences between the four populations was found ($\chi^2_6 = 121.27$, $p<0.05$), although there was no significant genotypic difference between the Australian European and Southern Chinese groups ($\chi^2_2 = 2.35$, $p>0.05$).

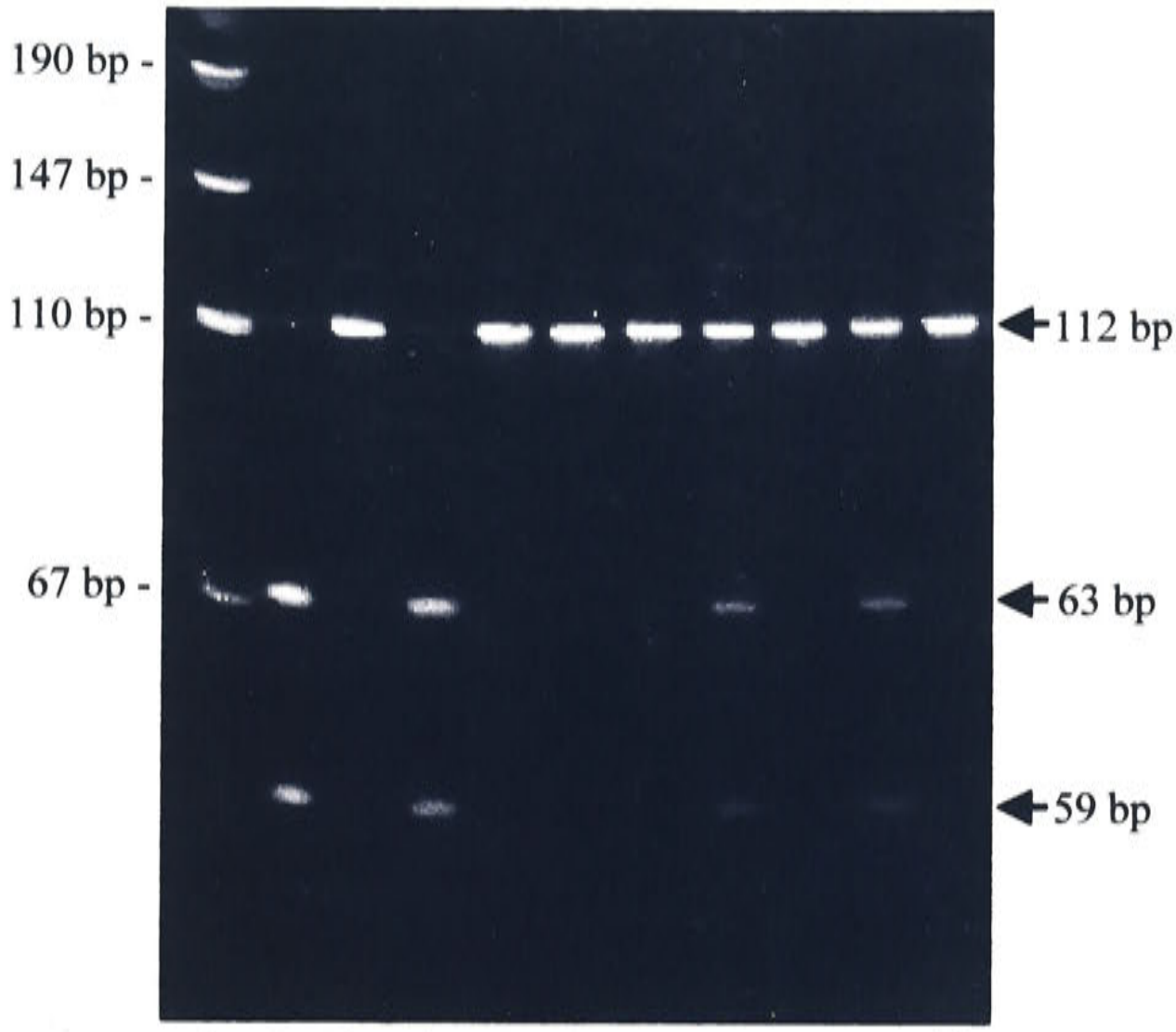
5.3.2 Detection of a 3 base pair Deletion Polymorphism

Three bands of unknown origin, in addition to the wild-type 127 bp fragment, were detected whilst genotyping samples for the p.A140D polymorphism in *GSTO1* exon 4. The four bands were observed in the three combinations described as follows: (1) the expected 127 bp wild-type pattern; (2) a 124 bp fragment and (3) a four-band pattern,



Fragment (bp)	C/C	C/A	A/A
127		—	—
68	—	—	
59	—	—	

Figure 5.3 - PCR/RFLP analysis of the *GSTO1* p.A140D variant. A 127 bp PCR product spanning exon 4 of the *GSTO1* gene was digested with *Cac*8I for nucleotide 419 determination. Lanes 1, 8 - C/A heterozygotes; Lanes 3-7 - C/C homozygotes; Lane 9 - A/A homozygote; Lane 2 - pUC/*Hpa*II marker.



Fragment (bp)	A/A	A/G	G/G
112	—	—	
63		—	—
59		—	—

Figure 5.4 - PCR/RFLP analysis of the *GSTO2* p.N142D variant. A 112 bp PCR product spanning exon 4 of the *GSTO2* gene was digested with *Sau*3AI for nucleotide 424 determination. Lanes 2, 4 - G/G homozygotes; Lanes 3, 5-7, 9, 11 - A/A homozygotes; Lanes 8, 10 - A/G heterozygotes; Lane 1 - pUC/*Hpa*II marker.

Table 5.7 – *GSTO1* and *GSTO2* allele and haplotype frequencies in three ethnic populations.

Gene	Polymorphism	Population	n	Genotype			Allele Frequency	
GSTO1	p.A140D	Australian	100	C/C=45	C/A=43	A/A=12	p.A140=0.665	p.D140=0.335
		Bantu African	72	C/C=58	C/A=14	A/A=0	p.A140=0.90	p.D140=0.10
		Creole African	81	C/C=55	C/A=23	A/A=3	p.A140=0.82	p.D140=0.18
		Chinese	100	C/C=71	C/A=25	A/A=4	p.A140=0.835	p.D140=0.165
	p.E155del	Australian	100	A*/A*=94	A*/B*=6	B*/B*=0	p.E155=0.97	p.del155=0.03
		Bantu African	68	A*/A*=66	A*/B*=2	B*/B*=0	p.E155=0.99	p.del155=0.01
		Creole African	85	A*/A*=75	A*/B*=10	B*/B*=0	p.E155=0.94	p.del155=0.06
		Chinese	100	A*/A*=91	A*/B*=8	B*/B*=1	p.E155=0.95	p.del155=0.05
GSTO2	p.N142D	Australian	100	A/A=49	A/G=40	G/G=11	p.N142=0.69	p.D142=0.31
		Bantu African	73	A/A=3	A/G=19	G/G=51	p.N142=0.17	p.D142=0.83
		Creole African	85	A/A=23	A/G=40	G/G=22	p.N142=0.51	p.D142=0.49
		Chinese	98	A/A=50	A/G=43	G/G=5	p.N142=0.73	p.D142=0.27

Gene	Population	n	Haplotype		
			*GSTO1*A* [p.A140;E155]	*GSTO1*B* [p.A140;del155]	*GSTO1*C* [p.D140;E155]
GSTO1	Australian	100	0.635	0.030	0.335
	Bantu African	68	0.900	0.015	0.081
	Creole African	81	0.760	0.060	0.180
	Chinese	100	0.785	0.050	0.165

Note: All polymorphisms are in agreement with the Hardy-Weinberg equilibrium.

with fragments of 124 and 127 bp and a doublet of approximately 130 and 131 bp (Figure 5.2). These bands persisted in the same samples with the use of two different sets of *GSTO1* Exon 4 primers (O1Ex4F and O1Ex4Rev2, and O1Ex4A2 and O1Ex4R2 in addition to O1Ex4F and O1Ex4Rev) and upon digestion with various restriction enzymes, indicating that the bands were not PCR artefacts, but due to the presence of a polymorphism undetected by database analysis.

One of the samples displaying the four-banded pattern was cloned into the pGEM®-T vector following the manufacturers protocol (Promega). The inserted DNA from the resulting clones was amplified by PCR and the products separated on 7% acrylamide gels. Clones containing either the 124 bp or 127 bp inserts were sequenced and revealed the deletion of three base pairs, AGg, spanning the exon4/intron4 boundary of *GSTO1* (Figures 5.5 and 5.6). Individuals with the 127 bp *GSTO1* exon 4 PCR products were found to be homozygous for the most common allele and termed *GSTO1*A*. Individuals with the 124 bp *GSTO1* exon 4 PCR products were found to be homozygous for the *GSTO1*B* allele. Samples displaying the four-band pattern were found to be heterozygous (*GSTO1*A*B*). When the two alleles combine, heteroduplexes are formed and give rise to the appearance of the two larger fragments. A similar occurrence has been reported in individuals heterozygous for the 3 bp deletion in *GSTM3* intron 6 [Inskip *et al.*, 1995].

There are two potential outcomes generated by the absence of the three bases at the exon 4/intron 4 boundary of *GSTO1*. Firstly, the disruption of the existing splice donor site may result in the transcription of a truncated protein due to the introduction of a premature stop codon in intron 4 (Figure 5.6A). Alternatively, deletion of AG from the p.E155 (GAG) codon, may allow the remaining G to form part of a new splice donor site (Figure 5.6B). The transcribed protein would lack the p.E155 residue. In order to determine the most likely outcome, *GSTO1* cDNA query sequences representative of each possible outcome were used to search the EST database. Query A represented the truncated version of *GSTO1*, and Query B the version lacking p.E155. The Query A sequence terminated 37 nucleotides early and ended with two nucleotides that are identical in both the intron 4 and exon 5 sequence (Figures 5.5 and 5.7A). 48 out of 50 ESTs (only 23/25 are represented) terminated at c.463, which corresponds to the third

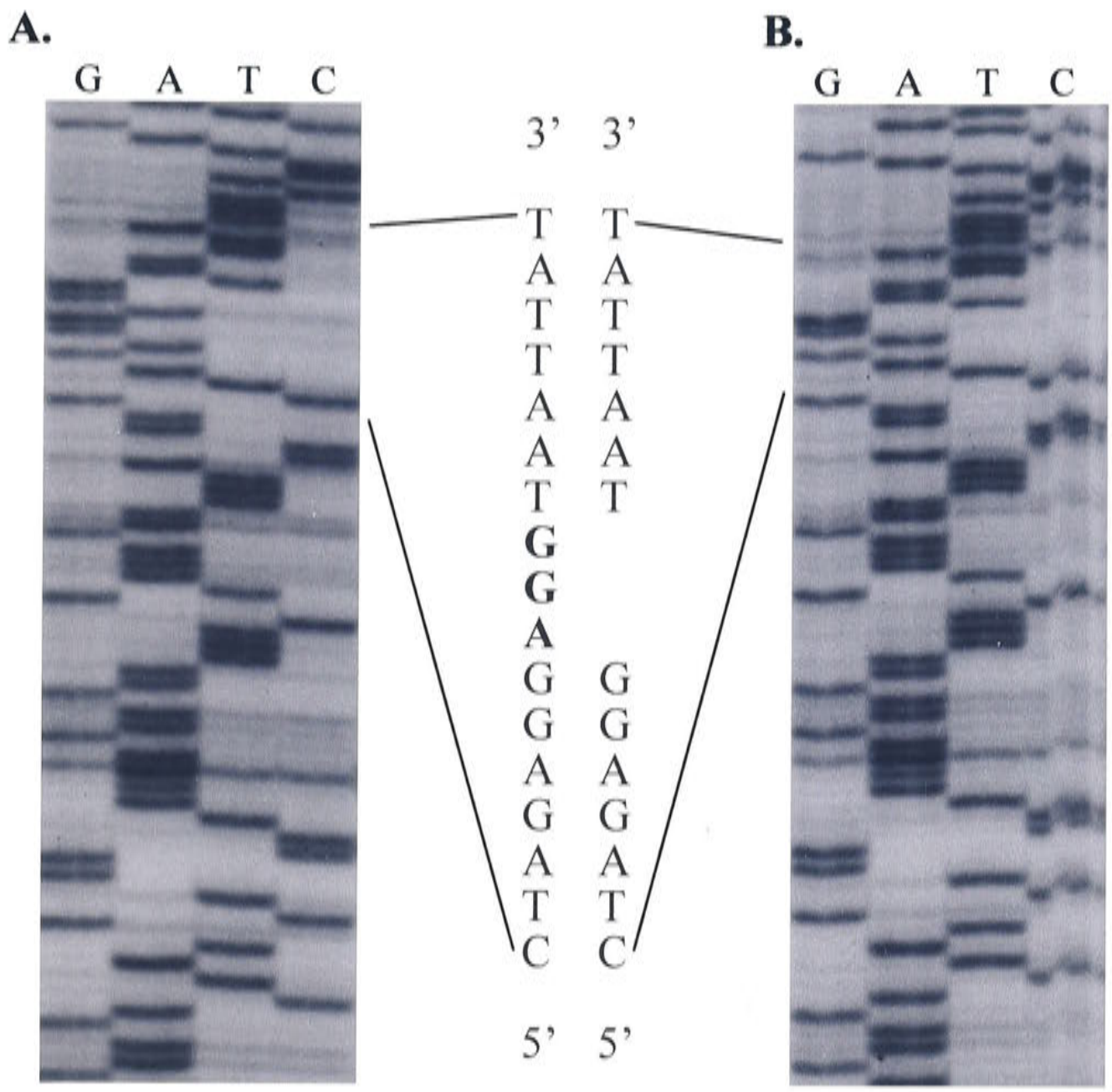


Figure 5.5 - Autoradiograph of a sequencing gel showing the *GSTO1*A* and *GSTO1*B* sequences. (A.) Wild-type *GSTO1*A* sequence. (B.) *GSTO1*B* sequence showing the deletions of three bases: 5'-AGG-3'.

A.

```
123  V   P   S   L   V   G   S   F   I   R   S   Q   N   K   E   D   Y   A
     GTGCCATCCTTGGTAGGAAGCTTTATTAGAAGCCAAAATAAAGAAGACTATGCT
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
     GTGCCATCCTTGGTAGGAAGCTTTATTAGAAGCCAAAATAAAGAAGACTATGCT
123  V   P   S   L   V   G   S   F   I   R   S   Q   N   K   E   D   Y   A

141  G   L   K   E   E   F   R   K   E   F   T   K   L   E   E
     GGCCTAAAAGAAGAATTTCGTAAAGAATTTACCAAGCTAGAGGAGgtaattatt
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
     GGCCTAAAAGAAGAATTTCGTAAAGAATTTACCAAGCTAGAGG---TAATTATT
141  G   L   K   E   E   F   R   K   E   F   T   K   L   E   V   I   I

     tctcctagctatcatcagagtaaacgataactatatctaccc  GSTO1*A
     |||||||||||||||||||||||||||||||||||||||||||
     TCTCCTAGCTATCATCAGAGTAAACGATAActatatctaccc  GSTO1*B
158  S   P   S   Y   H   Q   S   K   R   X
```

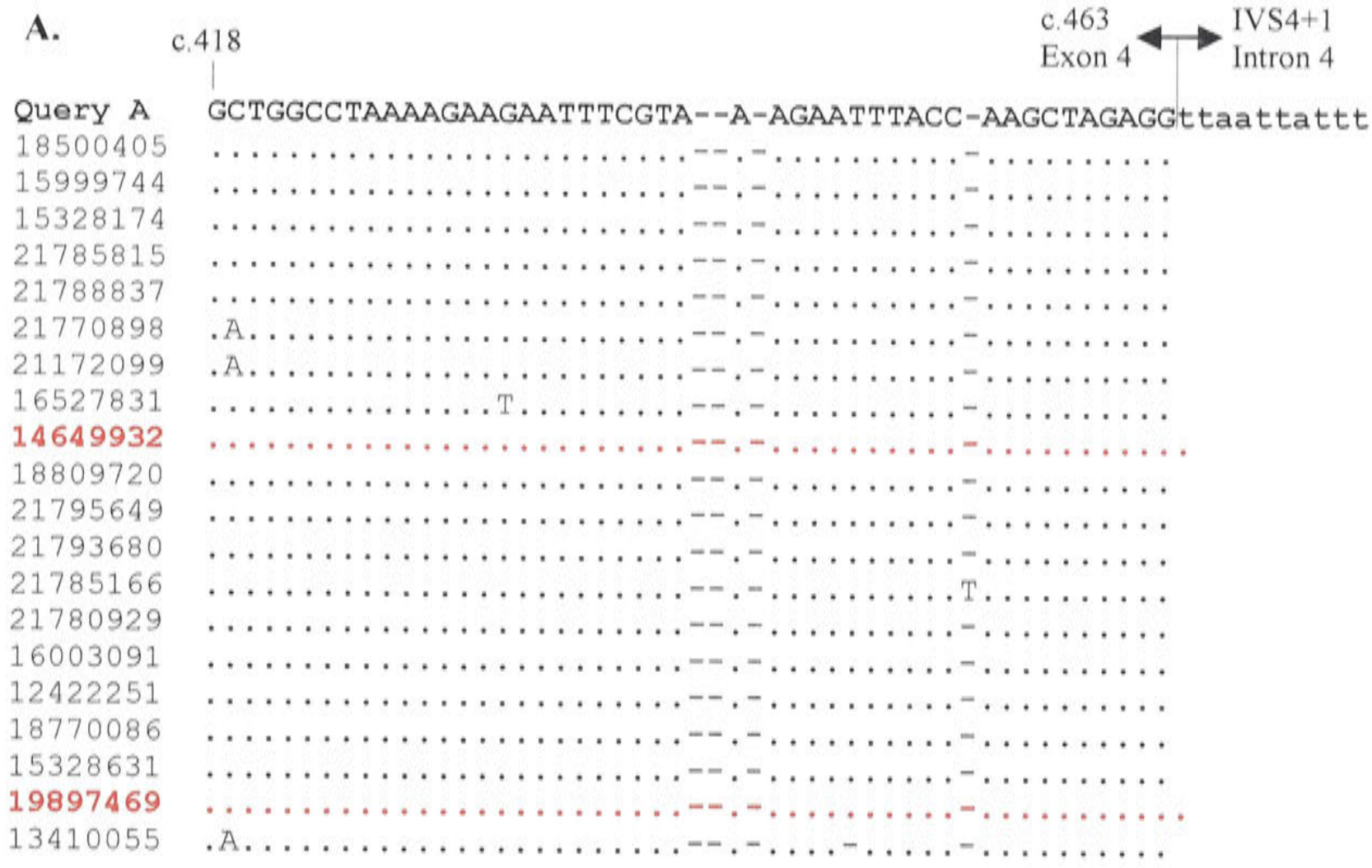
B.

```
149  E   F   T   K   L   E   E  Splice site
     GAATTTACCAAGCTAGAGGAGgtaattatttctcctagctat GSTO1*A
     ||||||||||||||||||||    ||||||||||||||||||||
     GAATTTACCAAGCTAGAGG---taattatttctcctagctat GSTO1*B
149  E   F   T   K   L   E  Splice site
```

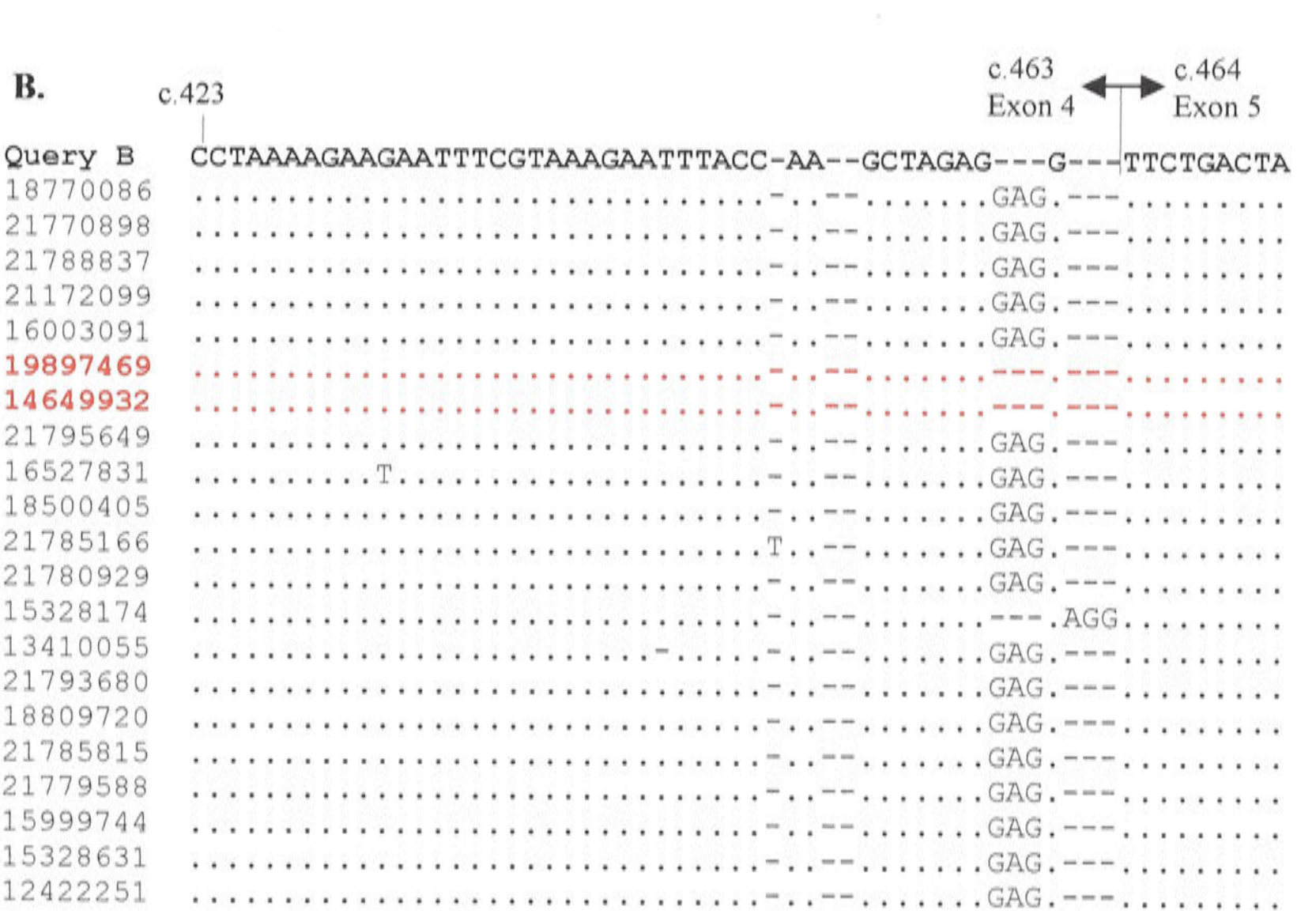
Figure 5.6 – The three base pair deletion spanning exon 4/intron 4 of *GSTO1* has two potential outcomes. Differences are highlighted in red. The exonic sequence is capitalised and the intronic sequence is shown in little letters. Residue names are centred above the middle nucleotide of the codon. (**A**). The 3 bp deletion may introduce a premature stop codon in intron 4. Comparison of exon 4 and part of intron 4 of the *GSTO1*A* and *GSTO1*B* sequences show the putative premature stop codon created in *GSTO1*B* due to the 3 bp deletion. (**B**). The 3 bp deletion may destroy the existing splice donor site. Comparison of exon 4 and part of intron 4 of the *GSTO1*A* and *GSTO1*B* sequences demonstrate that the splice donor site may reform using the first codon position G from p.E155, causing the deletion of p.E155.

Figure 5.7 – Representative alignment of human EST sequences with the two possible consequences of the *GSTO1* three base pair deletion used as the query sequence. Exonic sequences are capitalised, intronic sequences are represented by little letters. (A). Deletion of three bases across the exon4/intron 4 boundary of the *GSTO1* gene may disrupt the splice donor site hence the transcript will read through into intron 4 and create a premature stop codon, as represented by Query A. Most of the ESTs terminate at c.463, indicating that they do not read through into the intronic sequence. Two ESTs, highlighted in red, contain one extra nucleotide, which may represent part of the intron 4 sequence or part of exon 5. The fact that Query A sequence terminates 37 nucleotides early and ends with two nucleotides that are identical in both the intron 4 and exon 5 sequences indicates that no ESTs read through into the intron 4 sequence of *GSTO1*. (B). Deletion of three bases across the exon4/intron 4 boundary of the *GSTO1* gene may alternatively delete the glutamic acid encoded by residue 155. Only the two ESTs highlighted in red contain this deletion and continue with normal transcription.

A.


c.418
c.463
Exon 4
IVS4+1
Intron 4
Query A GCTGGCCTAAAAGAAGAATTTCGTA--A-AGAATTTACC-AAGCTAGAGGttaattattt
18500405 ..........--.-.........-.........
15999744 ..........--.-.........-.........
15328174 ..........--.-.........-.........
21785815 ..........--.-.........-.........
21788837 ..........--.-.........-.........
21770898 .A........--.-.........-.........
21172099 .A........--.-.........-.........
16527831 ..........T.--.-.........-.........
14649932 ..........--.-.........-.........
18809720 ..........--.-.........-.........
21795649 ..........--.-.........-.........
21793680 ..........--.-.........-.........
21785166 ..........--.-.........T.........
21780929 ..........--.-.........-.........
16003091 ..........--.-.........-.........
12422251 ..........--.-.........-.........
18770086 ..........--.-.........-.........
15328631 ..........--.-.........-.........
19897469 ..........--.-.........-.........
13410055 .A........--.-....-.....-.........

B.


c.423
c.463
Exon 4
c.464
Exon 5
Query B CCTAAAAGAAGAATTTCGTAAAGAATTTACC-AA--GCTAGAG---G---TTCTGACTA
18770086 ..........-..--.......GAG.---.........
21770898 ..........-..--.......GAG.---.........
21788837 ..........-..--.......GAG.---.........
21172099 ..........-..--.......GAG.---.........
16003091 ..........-..--.......GAG.---.........
19897469 ..........-..--.......---.---.........
14649932 ..........-..--.......---.---.........
21795649 ..........-..--.......GAG.---.........
16527831 .....T....-..--.......GAG.---.........
18500405 ..........-..--.......GAG.---.........
21785166 ..........T..--.......GAG.---.........
21780929 ..........-..--.......GAG.---.........
15328174 ..........-..--.......---.AGG.........
13410055 .....-....-..--.......GAG.---.........
21793680 ..........-..--.......GAG.---.........
18809720 ..........-..--.......GAG.---.........
21785815 ..........-..--.......GAG.---.........
21779588 ..........-..--.......GAG.---.........
15999744 ..........-..--.......GAG.---.........
15328631 ..........-..--.......GAG.---.........
12422251 ..........-..--.......GAG.---.........

position nucleotide in codon 154, indicating that they do not read through into the intronic sequence or exon 5. It was concluded that these ESTs did not contain the 3 base pair deletion. Two ESTs from different cDNA libraries, BI194912 and BQ068423, terminated at the same position as the Query A sequence. As these sequences all terminated 37 bases before the premature stop codon, it was assumed that these ESTs contained the three base pair deletion and that splicing continued as normal. This was confirmed by using Query B, which lacked p.E155, as the BLAST query sequence (Figure 5.7B).

As seen in Table 5.7, the *GSTO1*B* allele was relatively rare in the Australian European, Bantu African, Creole African and Southern Chinese populations studied here, with only one Southern Chinese individual homozygous for the *GSTO1*B* allele identified. No evidence for different genotypic distributions between each race was observed ($\chi^2_6 = 7.25$, $p>0.05$). The two allelic variants of *GSTO1* can be combined into four potential haplotypes, and these are described in Table 5.5. Only three of the four haplotypes were observed in the populations studied (Table 5.7).

5.3.3 Enzyme Characterisation

5.3.3.1 Structural Analysis

To evaluate the potential effects of the p.A140D and p.E155del polymorphisms may have on the GSTO1-1 protein, the substitutions were modelled into the 3-D structure of GSTO1-1 (Figure 5.8). Neither residue 140 or 155, both of which are located in the α5 helix, interact with the active site. The p.A140D change involves the substitution of alanine, a small nonpolar amino acid with the larger charged, polar aspartic acid, yet this substitution does not majorly affect the overall protein structure (Figure 5.9A), with no evidence of altered H-bond interactions detected (Figure 5.9B). In contrast, the deletion of residue 155 disrupts not only the structure of the helix, but also the entire protein (Figure 5.10A), causing substantial conformational changes to the catalytically essential p.C32 residue and the subsequent introduction of two new H-bonds between p.C32 and p.A35, and p.S28 and p.M29 (Figure 5.10B).

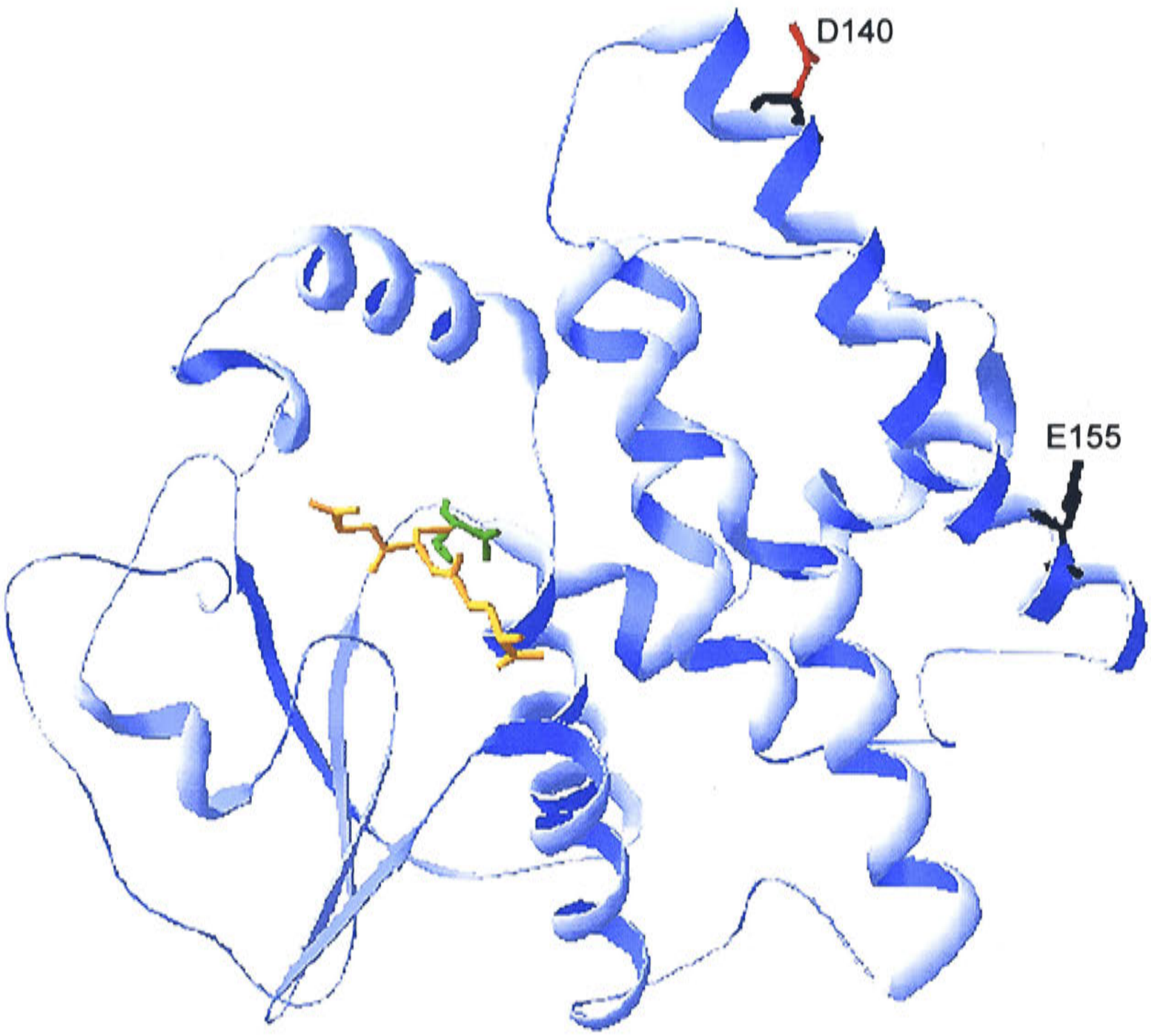
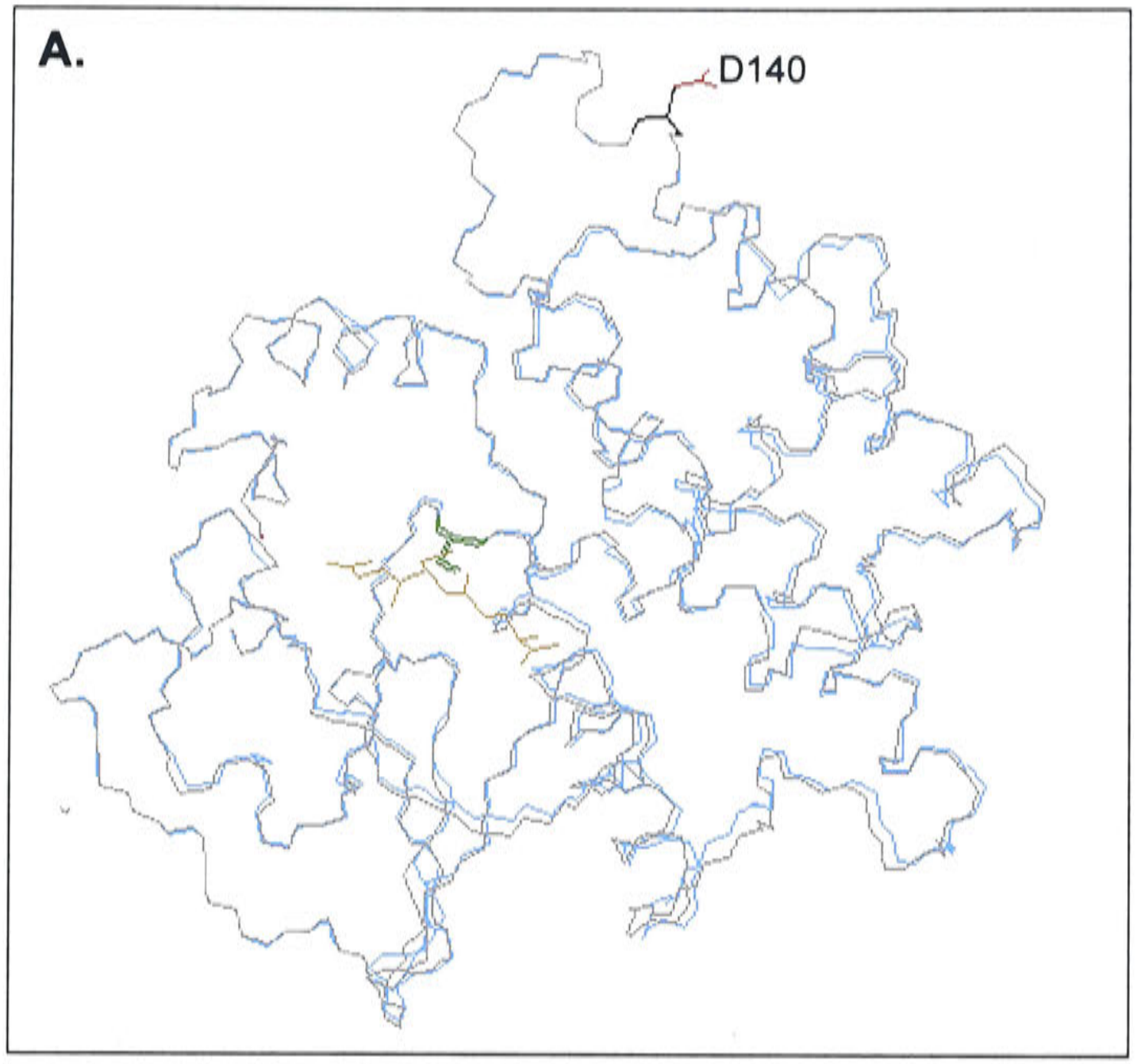


Figure 5.8 – Ribbon diagram of GSTO1 depicting the position of the polymorphic p.A140D substitution and the position of the deleted p.E155 residue. The most common amino acids, p.A140 and p.E155, are shown in black. The least common amino acid, p.D140, is coloured red. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.C32 residue is coloured green.

Figure 5.9 – Carbon backbone of the GSTO1 monomer encoding p.A140 overlaid with the protein encoding p.D140. GSTO1 encoding the more common amino acid p.A140 (depicted in black) is coloured grey; GSTO1 encoding the less common amino acid p.D140 (depicted in red) is coloured blue. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.C32 residue is coloured green. (A). Substitution of the p.D140 residue for the p.A140 residue does not alter the overall structure of the GSTO1 protein. (B). Magnification of residue 140 and neighbouring residues. H-bonds common to both variants are coloured dark blue; those introduced by the substitution of p.A140 for p.D140 are shown in pink.


A.
D140

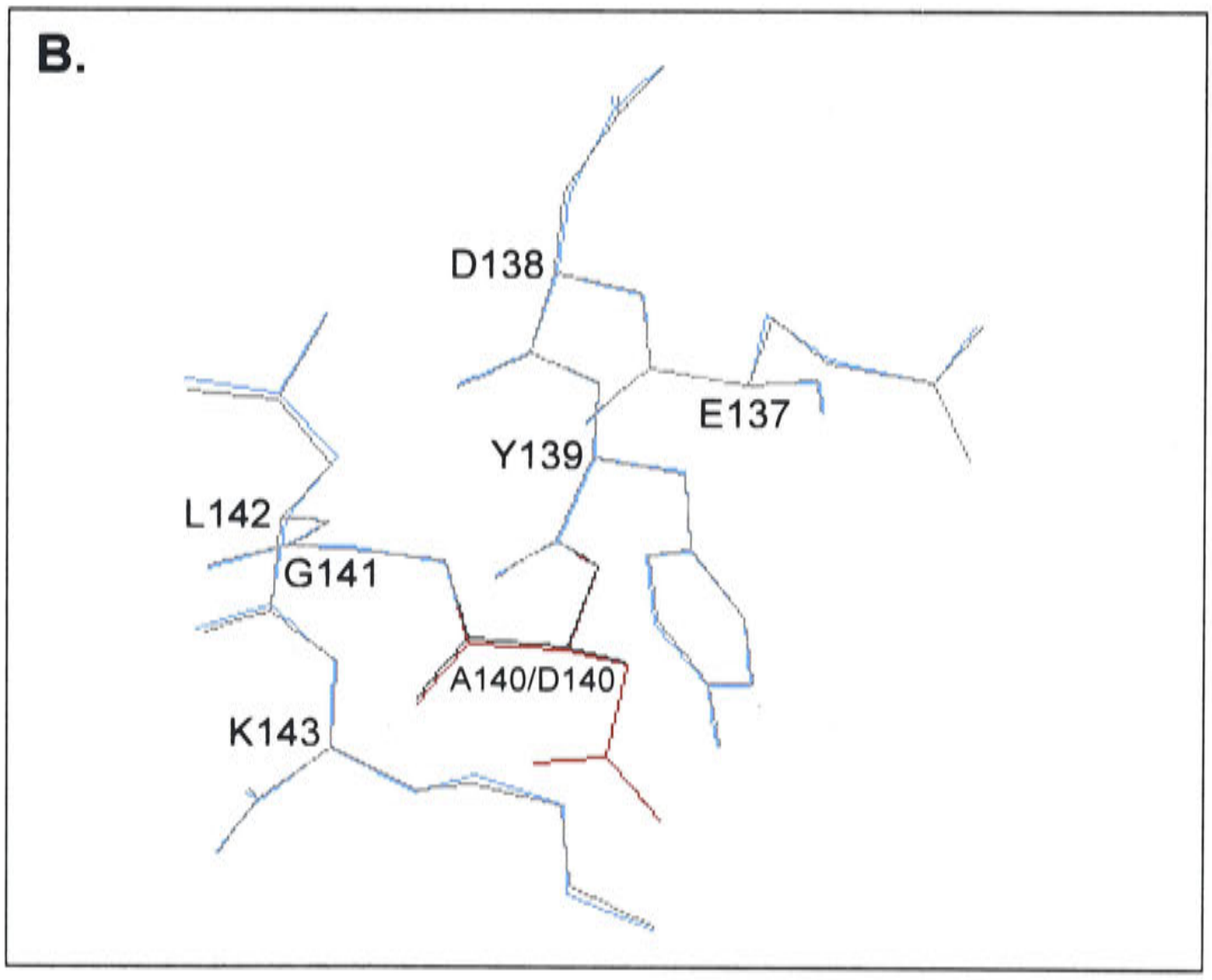
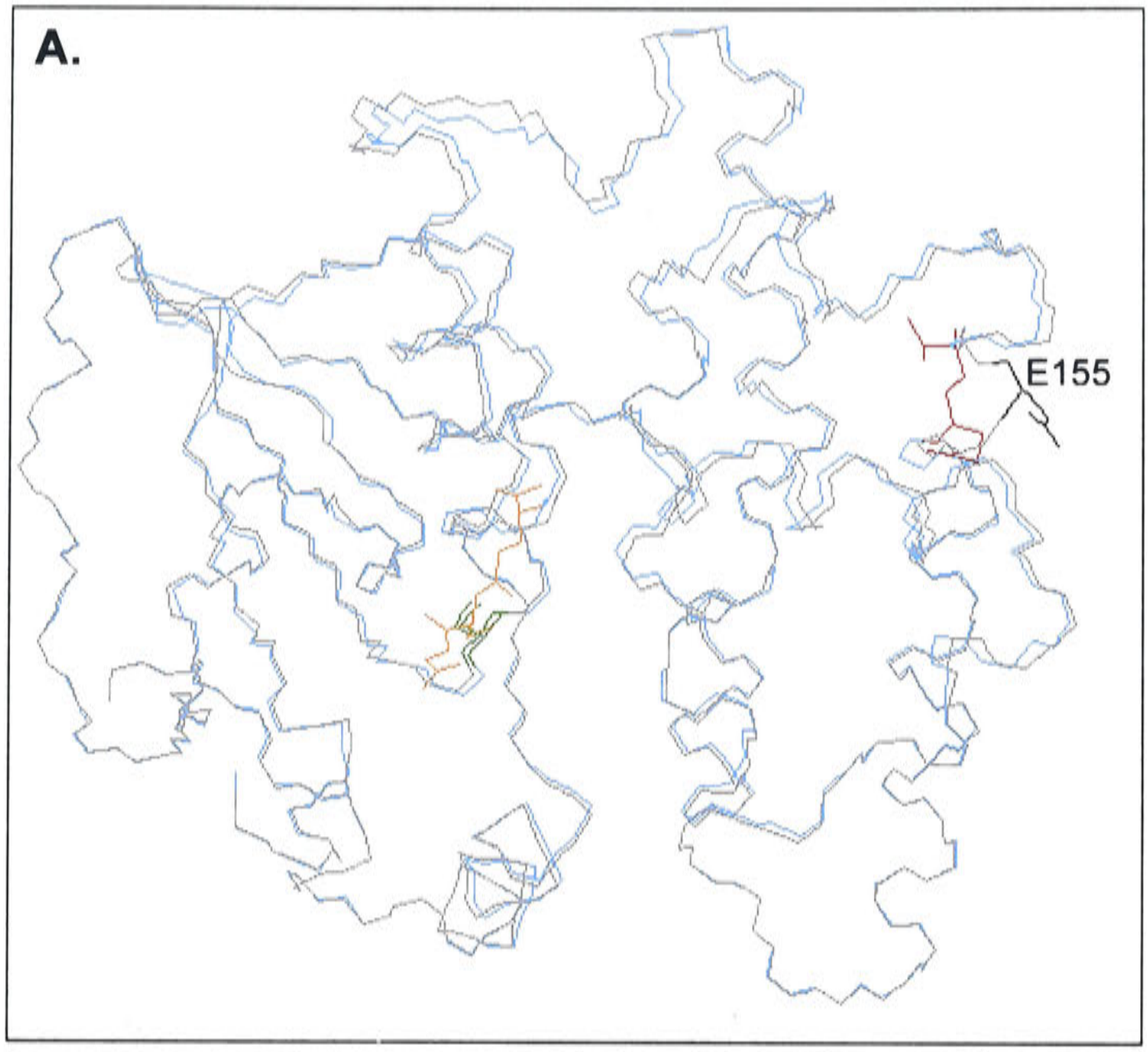
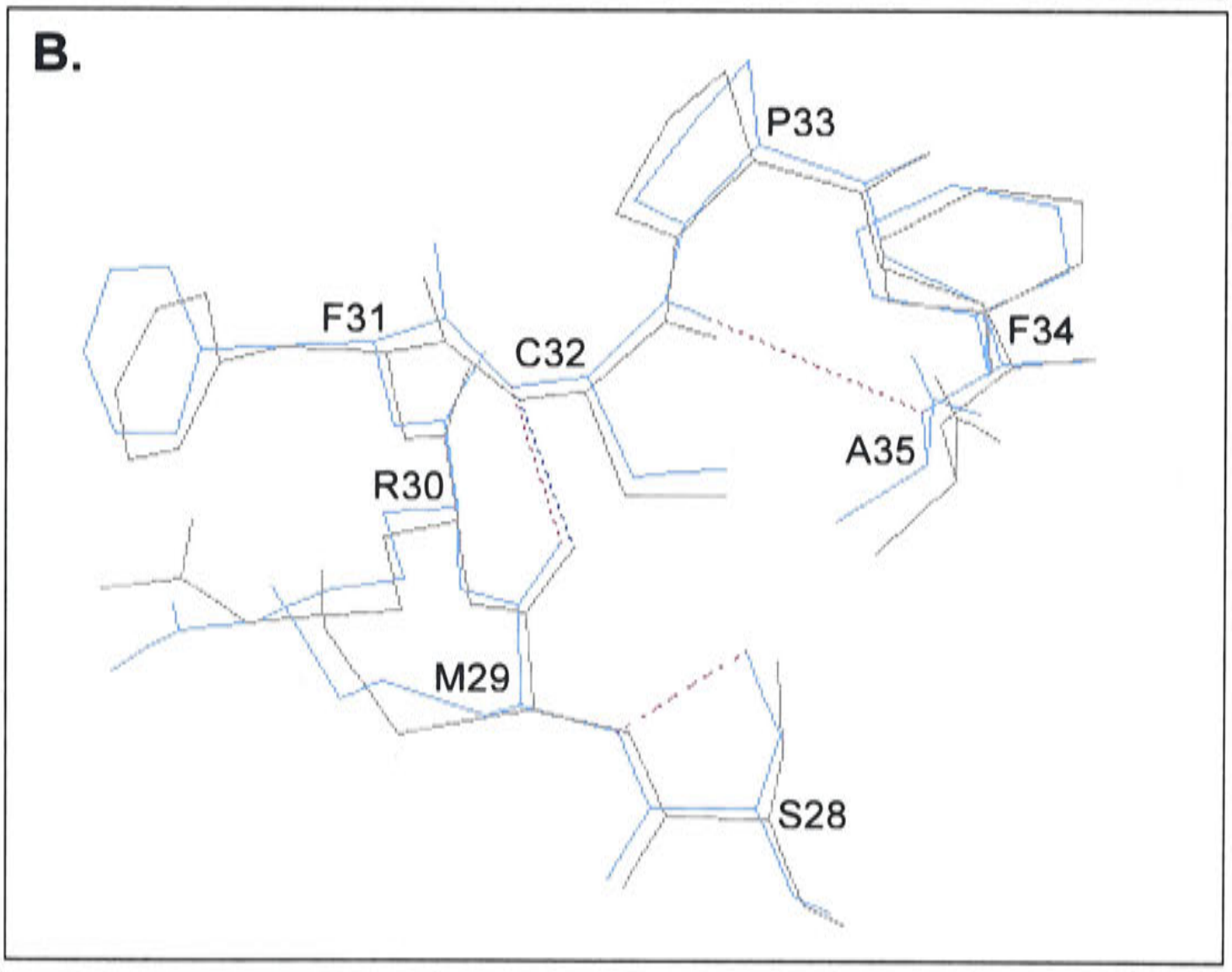

B.
D138
E137
Y139
L142
G141
A140/D140
K143

Figure 5.10 – Carbon backbone of the GSTO1 monomer encoding p.E155 overlaid with the protein missing p.E155. GSTO1 encoding p.E155 (depicted in black) is coloured grey; GSTO1 in which p.E155 is deleted is coloured blue, and the residues normally adjacent to p.E155 are depicted in red. Glutathione, shown in gold, is bound in the active site. The catalytically essential p.C32 residue is coloured green. (A). Deletion of the p.E155 residue introduces considerable alterations to the overall structure of the GSTO1 protein. (B). Magnification of residue 32 and neighbouring residues. H-bonds common to both variants are coloured dark blue; those introduced by the deletion of p.E155 are shown in pink.


A.
E155


B.
P33
F31
C32
F34
A35
R30
M29
S28

5.3.3.2 ENZYMATIC ANALYSIS

Site directed mutagenesis was used to create the GSTO1*A, GSTO1*B and GSTO1*C allelic recombinant isoforms in order to determine whether these variants had an effect on the catalytic efficiency of the protein. As predicted by the structural modelling, no significant differences in activity were observed between the p.A140 and p.D140 variants. However, significantly higher activities towards both CDNB and HEDS were measured for the GSTO1*B variant ($p<0.001$), with a 1.34-fold and 1.38-fold increase in specific activity respectively (Table 5.8).

The GSTO2*A and GSTO2*B recombinant allelic isoforms were unable to be characterised enzymatically due to protein insolubility problems.

Table 5.8 – Specific activities of recombinant GSTO1-1 variants towards the substrates CDNB and HEDS.

Substrate	**GSTO1*A [p.A140;E155]**	**GSTO1*B [p.A140;del155]**	**GSTO1*C [p.D140;E155]**
		µmol/min/mg	
1-chloro-2,4-dinitrobenzene	0.047 ± 0.014	0.063 ± 0.011	0.039 ± 0.013
2-hydroxyethyl disulfide	2.616 ± 0.262	3.604 ± 0.364	2.499 ± 0.177

Results are the mean ± S.D. for at least ten determinations.

5.4 DISCUSSION

Despite the lack of detoxification activities currently ascribed to the recently discovered Omega class GSTs, a number of novel physiological roles never before described in the GST superfamily have been attributed to GSTO1-1. With its broad expression profile, weak thiol transferase and dehydroascorbate reductase activities, involvement in the biomethylation of inorganic arsenic and regulation of ryanodine receptor activity, it would appear that GSTO1-1 has a fundamental role in cellular metabolism [Board *et al.*, 2000]. Polymorphisms in the *GSTO1* gene could potentially alter these activities with unknown consequences. By discovering novel polymorphisms in the *GSTO1* gene, the roles underlying these processes can be further investigated. Unfortunately, although polymorphisms can be identified in the *GSTO2* gene, the function of this gene, and therefore any polymorphic consequences, remain unknown due to protein expression problems. This is currently being addressed in this laboratory.

Polymorphisms in the Omega class GSTs were detected using an approach combining the use of computer based programs to search the EST and UniGene sequence databases, and the analysis of SNP databases for the presence of GST Omega variants. Variants were discovered in both the *GSTO1* and *GSTO2* genes by using the BLAST alignment tool to search the EST database, although there was a distinct lack of variation detected in *GSTO2* (Table 5.1) possibly due to poor representation in the EST database. In contrast to *GSTO1*, which was well represented by 100 ESTs, *GSTO2* was moderately represented by 46 ESTs. As GSTO2-2 is moderately expressed in a range of tissues [Whitbread *et al.*, In Press], the existence of other polymorphisms in this gene is unlikely to be detected until more full-length *GSTO2* cDNA sequences are submitted to the EST database.

Polymorphisms in the *GSTO1* gene were discovered using the SNP Finder program to search the *GSTO1* UniGene cluster, which consisted of 179 EST and mRNA sequences. Analysis of *GSTO2* could not proceed due to the lack of a representative *GSTO2* UniGene cluster (Table 5.2). UniGene clusters are designed using information pertaining to characterised genes, with each cluster composed of ESTs and full-length mRNA sequences that have been deposited in various sequence databases. Sequences

that cannot be clustered are considered novel [Boguski & Schuler, 1995; Schuler *et al.*, 1996]. The *GSTO2* gene, which has only recently been identified [Whitbread *et al.*, In Press], falls into this category.

Analysis of the SNP databases for the presence of GST Omega sequences revealed the presence of *GSTO1* (Table 5.2) in all but the EGSNPs and GENESNPs databases. The SNPper SNP detection program was also unable to detect any *GSTO1* variants in the main SNP databases. In contrast, *GSTO2* was not represented in any of the SNP databases analysed. One possible explanation for the absence of *GSTO2* variants could be that these had been mistakenly identified as *GSTO1* variants and hence listed with other *GSTO1* submissions. However, this is unlikely as the *GSTO1* and *GSTO2* sequences share only 64% sequence identity [Whitbread *et al.*, In Press], and there was no evidence that this had happened. Therefore, as discussed above, the absence of *GSTO2* is most likely to be due to its novel status.

Using this combined database mining approach, a total of ten novel missense variants were identified and verified in the *GSTO1* and *GSTO2* genes: seven were identified using the BLAST alignment tool and SNP Finder program, and three were detected in the SNP databases (Tables 5.1 and 5.2). Of these, only two were present in the Australian European, Bantu African, Creole African and Southern Chinese populations – the *GSTO1* p.A140D and the *GSTO2* p.N142D polymorphisms. The remaining variants may be rare alleles, and it is most likely that the absence of the *GSTO1* p.S121P, p.K136X and p.E154D polymorphisms in these populations fall into this category. These three variants were identified by CGAP with the use of the SNP Finder program and were subsequently submitted to dbSNP. All three variants have an extremely low likelihood score (Table 5.2), which generally implies that these variants may not be real. Small-scale population screening was continued as analysis of the original trace data of the relevant sequences confirmed that these sequence alterations were real. Although only one sequence was verified for the *GSTO1* p.S121P and p.K136X variants, three cDNAs generated from two different libraries were confirmed for the p.E154D variant. The absence of the remaining *GSTO1* variants, especially those with likelihood scores greater than 0.9 (Table 5.2) may be due to associations with ethnic groups not screened in the present study, as the origin of the cDNA in the EST and UniGene databases is unknown.

The only confirmed *GSTO2* polymorphism detected using this bioinformatic approach, the p.N142D polymorphism, was also detected during independent cloning of a *GSTO2* cDNA for characterisation studies of the *GSTO2* gene [Whitbread *et al.*, In Press]. This polymorphism was found to be quite common in the populations screened in this study. The p.D142 allele occurred most frequently in the two African populations. This was found in approximately 50% of the Creole population screened, and at 83% appeared to be the common allele in the Bantu African population. The Australian European and Chinese groups showed similar allele distributions, with the p.N142 variant being the most common allele. Unfortunately, due to *GSTO2* protein purification problems, subsequent enzymatic characterisation of this polymorphism was unable to be investigated in the course of this study. However, it can be hypothesised that although this polymorphism does not occur in the active sites of GSTO2-2, substitution of the uncharged, polar asparagine residue for the charged aspartic acid residue, which is also polar and of a similar size, may affect the structural integrity of the protein, perhaps due to indirect interactions with active site residues, packing or backbone stabilisation.

The *GSTO1* p.A140D polymorphism was also relatively common in all the populations tested, although the frequency of the p.D140 allele was notably lower in the Bantu African group at 10%. No significant alterations in thiol-transferase or GST activity were detected during enzymatic characterisation of the GSTO1*A and GSTO1*C allelic isoforms. Although the p.A140D polymorphism involves the non-conservative substitution of alanine, a small nonpolar amino acid with the larger charged, polar aspartic acid, the external orientation of residue 140 at the N-terminal end of helix α5 makes it unlikely that this will influence GSTO1-1 activity. These results are consistent with predictions based on 3-D modelling, which demonstrated that the A140D substitution had no effect on the structure of the protein (Figure 5.9). However, with the exception of MMA^V, which could not be tested in this study, a unique GSTO1-1 substrate has not been identified, hence the full effects of the GSTO1-1 p.A140D polymorphism will remain unknown until substrates specific for GSTO1-1 are identified.

During the *GSTO1* polymorphism screening process, a novel three base pair deletion occurring in a trinucleotide repeat was detected spanning the exon 4/intron 4 boundary of *GSTO1* (Figure 5.6). The majority of deletion polymorphisms occur within repetitive

regions with trimeric and tetrameric short tandem repeats, prime targets for DNA polymerase slippage, being especially polymorphic within the human genome [Edwards *et al.*, 1989; Edwards *et al.*, 1991; Krawczak & Cooper, 1991]. The deletion of AG from GAG, the last codon in exon 4 (p.E155), allowed the G in the first codon position to replace the deleted splice donor site G (Figure 5.6B). This in effect deleted the p.E155 residue and created a new splice donor site. Splicing was predicted to continue as usual, as the splice donor site consensus sequence containing the invariant GT dinucleotide [Mount, 1982] was not altered. This polymorphism was found to be relatively rare in the three populations genotyped in this study, with only one homozygous individual detected in the Southern Chinese population. Enzymatic characterisation of the GSTO1*B recombinant allelic isoform demonstrated a significant increase in enzyme activity when compared with the GSTO1*A protein. Although residue 155 is not located in the active sites, it is possible that its location near the C-terminal end of helix α5 may be important for maintaining the structure and stability of the helix, hence deletion of this residue may disrupt protein stability and alter protein function. Conformational changes to the overall structure of the protein were supported by predictive modelling (Figure 5.10A). In addition, substantial conformational changes were introduced around the catalytically essential p.C32 residue, causing the creation of three new H-bonds (Figure 5.10B). It is possible that the combined effects of these changes are responsible for the observed alterations in enzymatic activity.

Only one other small deletion polymorphism has been described in the GST superfamily – the three base pair deletion occurring in intron 6 of the *GSTM3* gene [Inskip *et al.*, 1995]. This *GSTM3* polymorphism has been associated with altered expression levels of GSTM3-3, and is strongly associated with another known Mu class polymorphism encoded by the *GSTM1*A* allele. Together these have been implicated in moderating susceptibility to various diseases [Yengi *et al.*, 1996; Jahnke *et al.*, 1996]. It is possible that the *GSTO1*A* or *GSTO1*B* alleles could also be linked with other Omega class polymorphisms or with other loci. Further investigations are warranted.

GSTO1 has recently been identified as MMA^V reductase, an enzyme involved in the bioactivation pathway of inorganic arsenic (Figure 5.1) [Zakharyan *et al.*, 2001]. It is therefore possible that *GSTO1* polymorphisms could alter this activity. GSTO1-1

catalyses the reduction of MMA^{V} to MMA^{III}, which is more toxic than the parent compound arsenite in mammalian tissues, suggesting that this may be responsible for the pathological effects usually attributed to arsenic [Petrick *et al.*, 2000; Sampayo-Reyes *et al.*, 2000]. As many tissues have the capacity to produce the highly toxic intermediate MMA^{III}, due to the wide tissue distribution of GSTO1-1 [Board *et al.*, 2000] and the other enzymes involved in this pathway, each may be a potential target organ for arsenic toxicity and/or carcinogenicity [Sampayo-Reyes *et al.*, 2000]. Many studies have reported correlations between high levels of arsenic in drinking water sources with various arsenic associated diseases such as cancers of the skin and various internal organs [Chen *et al.*, 1985], Blackfoot disease in Taiwanese populations [Tseng, 1977; Chen *et al.*, 1985], cutaneous disorders in rural villages in North Mexico [Cebrian *et al.*, 1983] and open lesions in West Bengal [Guha Mazumder *et al.*, 1988].

Despite the toxicity of arsenic, tolerance to arsenic poisoning has been described in populations exposed to high levels of inorganic arsenic. Consumption of lethal doses of arsenic with no apparent ill effect was first reported in 1862 in the Styrian population located in the Austrian Alps [Roscoe, 1862; Aposhian, 1997]. Subsequent studies have concentrated on areas where the drinking water is contaminated with levels of inorganic arsenic exceeding the 10 μgAs/L recommended by WHO [Yamamura *et al.*, 2001]. Residents in the Chilean village of San Pedro de Atacama were found to have a higher body burden of arsenic, but did not seem to develop arsenic associated disease despite the extremely high (600 μgAs/L) levels of arsenic contaminating their drinking water [Aposhian *et al.*, 1997]. Large variations in the urinary biomarkers used to measure arsenic exposure (MMA, DMA and inorganic arsenic) have also been reported in Andean women from four different northern Argentinian village populations [Vahter *et al.*, 1995], Romanians, Chileans and Mongolians [Aposhian *et al.*, 2000a; Aposhian *et al.*, 2000b]. It has been suggested that arsenic tolerance and these variations may be associated with polymorphisms in the enzymes involved in the bioactivation pathway, especially in the rate-limiting MMA^{V} reductase GSTO1-1 [Vahter *et al.*, 1995; Aposhian *et al.*, 1997; Zakharyan & Aposhian, 1999; Aposhian *et al.*, 2000a; Aposhian *et al.*, 2000b]. It is possible that at least one of the *GSTO1* polymorphisms described in the present study may be associated with these observed variations. It would be of interest to genotype populations exposed to high levels of inorganic arsenic. These

populations could be further grouped into (1) those that develop arsenic associated disease, (2) those that do not develop arsenic associated disease and (3) those with variable urinary concentrations of MMA, DMA and inorganic arsenic.

The role of GSTO1-1 in the metabolism of inorganic arsenic may also extend to the application of inorganic arsenic in chemotherapy. Inorganic arsenic is currently used as a chemotherapeutic agent for the treatment of many cancers, in particular AML. It is thought to act by causing partial cytodifferentiation of tumour cells, caspase activation and subsequent induction of apoptosis [Soignet *et al.*, 1998]. Although this is an effective treatment, with only mild to moderate side effects usually reported [Shen *et al.*, 1997; Ohnishi *et al.*, 2000; Soignet *et al.*, 2001], concern over its use has increased with the recent unexpected deaths of three AML patients receiving arsenic trioxide treatment [Westervelt *et al.*, 2001]. Notably, these patients were all of African American descent. It is believed that the different susceptibility to the toxicity of arsenic trioxide, as manifested by varied side effects, may be due to polymorphisms in the genes responsible for arsenic metabolism [Westervelt *et al.*, 2001]. It is possible that the polymorphisms identified in this study may also be associated with this variation, in particular the p.A140D polymorphism. As seen in Table 5.7, homozygotes for the p.D140 allele were severely under-represented in the Creole African population, and absent in the Bantu African population. It is therefore important to determine whether this polymorphism, in addition to the deletion polymorphism identified in this study, do in fact influence the biomethylation of inorganic arsenic in order that AML and other cancer patients carrying these polymorphisms can be administered a safer chemotherapeutic treatment.

Polymorphisms in GSTO1-1 may also be of importance in patients with radiation resistant tumours. An association between the over-expression of the mouse GSTO1-1 orthologue p28 with the observed resistance of mouse lymphoma cell lines to radiation and chemotherapeutics [Kodym *et al.*, 1999]. Over-expression of various classes of GSTs has been associated with resistance to different chemotherapeutic agents [Tew, 1994], hence it is possible that over-expression of the GSTO1-1 protein may lead to tumour cell resistance to inorganic arsenic in addition to resistance to radiation. As mutations are known to alter drug resistance [Hayes & Wolf, 1990], the polymorphisms in GSTO1-1 described in this study could hypothetically alter not only the

chemotherapeutic effects of inorganic arsenic, but also the effects of radiation in a wide range of target tissues.

The two *GSTO1* polymorphisms may also affect GSTO1-1 modulation of ryanodine receptors RyR1 and RyR2, thereby regulating the release of Ca^{2+} from intracellular stores. RyR1 and RyR2 have been detected in various tissues [Xu *et al.*, 1998a] in addition to the skeletal and cardiac muscle tissue in which RyR1 and RyR2 respectively are typically expressed [Takeshima *et al.*, 1989; Otsu *et al.*, 1990; Zorzato *et al.*, 1990]. RyR1 has also been detected in B-lymphocytes, which require Ca^{2+} for B-cell response determination [Sei *et al.*, 1999]. Activation of RyRs increases cytoplasmic Ca^{2+} levels during oxidative stress [Dulhunty *et al.*, 2000]. As an increased cytoplasmic Ca^{2+} concentration is an essential stage during the induction of apoptosis, RyRs activation could be responsible for triggering apoptosis in both normal and tumour cells [Berridge *et al.*, 1998; Mariot *et al.*, 2000; Pan *et al.*, 2000]. The finding that GSTO1-1 is able to potentiate RyR1 and inhibit RyR2 has also led to the suggestion that it may be involved in the modulation of B-cell immune responses in addition to conferring protection against apoptosis and oxidative stress through inhibition of RyR2 activation [Dulhunty *et al.*, 2001]. Polymorphisms in GSTO1-1 could possibly alter these protective effects.

As discussed in Chapters 3 and 4, there are three important points to consider when assessing the effects polymorphisms may have on protein expression and function. Firstly the effects of interindividual variation of protein expression must be taken into account as it has been suggested that the resultant variation in enzyme activity may be greater than that caused by genetic polymorphisms [Coles *et al.*, 2000]. Although interindividual variation of GSTO1-1 expression has not yet been reported, the marked variation observed for many members of the GST superfamily [Hayes *et al.*, 1989; Singhal *et al.* 1993; Anttila *et al.*, 1995; Nakajima *et al.*, 1995; Mulder *et al.*, 1999; Coles *et al.*, 2000] suggests that this should not be overlooked. Secondly, linkage disequilibrium between polymorphisms in both the Alpha and Mu class GSTs been shown to affect the expression levels of GSTA1-1, GSTA2-2 and GSTM3-3 [Inskip *et al.*, 1995; Coles *et al.*, 2001a]. The *GSTO1* p.E155del polymorphism in particular may also exert a similar effect through tissue-specific alterations in splicing efficiency. Finally, small but significant differences in GST Omega allele distributions were observed between the four ethnic populations tested here. As variation between

populations is influenced by differences in genetic, environmental or dietary interactions [Kawajiri *et al.*, 1990; Garte, 1998; Park *et al.*, 2000; Risch, 2000], the effect one particular GSTO1-1 polymorphism may have on protein function in one population will not necessarily be the same as the effect the same polymorphism has in another population.

Using a combination of database mining strategies, two novel polymorphisms were identified in the human Omega class GST genes. Although each method yielded results for the *GSTO1* gene, polymorphisms in the newly characterised *GSTO2* gene were only found in the EST database. The discovery of a novel three base pair deletion in the *GSTO1* gene through laboratory-based methods emphasises the point that identification of polymorphisms should not be limited to one method or technique. Functional and structural differences were observed for the *GSTO1* polymorphisms using the substrates currently available, however it is not currently known whether these polymorphisms have any clinical implications. The identification of other GSTO1-1 specific substrates will help determine whether these polymorphisms exert any physiological effects.

CHAPTER 6

FINAL CONCLUSIONS

Research into SNP detection has rapidly expanded in recent years largely due to the recent completion of the Human Genome Project, the advent of sequence databases and the recognition that some polymorphisms have physiological consequences in disease susceptibility and adverse drug reactions. Facilitating this research is the development of constantly evolving SNP databases and computer-based tools designed to detect SNPs within genes of interest. These recent advances in SNP detection were used to identify novel polymorphisms in the Alpha, Mu and Omega class GSTs, which comprise part of the GST superfamily of xenobiotic metabolising enzymes. Overall, the combined use of various SNP databases, and the BLAST and SNP Finder computer algorithms to search the EST and UniGene databases respectively, has proved an efficient, cost-effective and rapid method of discovering novel coding region polymorphisms in the GST genes.

No single program detected all of the polymorphisms described in this study. However, at least one novel polymorphism was isolated by each method. This suggests that the application of more than one SNP detection method should be used if a gene is to be comprehensively screened for polymorphisms present in the public sequence and SNP databases. The lack of overlapping entries between databases has also been emphasised in a recent study in which it was reported that when a search was limited to three SNP databases, only eight of the 893 SNPs the group had previously detected were common to all three selected databases [Marsh *et al.*, 2002].

Using a combined approach, a total of eight uncharacterised polymorphisms were identified in this study. These include the *GSTA2* p.P110S, p.T112S and p.E210A; the *GSTA3* p.L71I; the *GSTM3* p.W147G and p.V224I; the *GSTO1* p.A140D and the *GSTO2* p.N142D polymorphisms. In addition, genotyping assays of putative *GSTO1* polymorphisms subsequently revealed the presence of a novel three base pair deletion in the *GSTO1* gene. Of these, the *GSTA2* p.P110S, the *GSTM3* p.W147G and p.V224I, and the *GSTO1* p.E155del polymorphisms were found to have significant effects upon the protein function. The Pi, Theta 1 and Zeta class GSTs were also analysed for the existence of novel polymorphisms, although polymorphism in both the *GSTP1* and *GSTZ1* genes has been investigated extensively [Ali-Osman *et al.*, 1997; Blackburn *et al.*, 2000; Blackburn *et al.*, 2001]. Preliminary population genotyping studies have

confirmed the presence of two novel polymorphisms in the Theta and Zeta class GSTs – the *GSTT1* p.V95L and the *GSTZ1* p.N133H polymorphisms (Appendix 4).

A number of advantages and disadvantages were encountered with the application of database mining techniques, especially during the detection of polymorphisms present in the public sequence databases. The greatest limitation is the high yield of false positive results, as observed in this study. The combined results from Chapters 3, 4 and 5 demonstrate that only 35% of the Alpha, Mu and Omega class GST missense polymorphisms that were detected using the *in silico* BLAST Alignment tool and SNP Finder methods and subsequently verified, and only 18% of the missense polymorphisms identified in the SNP databases, could be confirmed in the four populations screened in this study. Other groups mining the EST and UniGene databases for SNP information have reported similar results despite the use of different computer algorithms designed to search these databases and the implementation of stringent filtering processes. The number of seemingly genuine polymorphisms confirmed in subsequent population studies has been found to range anywhere between 20%-82% [Gu *et al.*, 1998; Buetow *et al.*, 1999; Forsberg *et al.*, 1999; Garg *et al.*, 1999; Picoult-Newberg *et al.*, 1999; Irizarry *et al.*, 2000].

If the success rate of SNP detection using methods based on database analysis is to improve, the source of the problems responsible for the high rate of false positives must be addressed. In respect to the EST and UniGene databases, the high error rate of the EST information contained within these databases partially contributes to this problem. The combination of single pass, high throughput automated sequencing of cDNA libraries of variable quality, and the use of error prone reverse transcriptase used to generate the ESTs automatically introduces an inherent 2%-3% error into each EST [Adams *et al.*, 1991; Gu *et al.*, 1998]. In addition, some cDNA libraries are constructed from tumour tissues, which could lead to the detection of spontaneous tumorigenesis-associated polymorphisms [Gu *et al.*, 1998]. Even though these spontaneous mutations can and will be validated in the relevant cDNA clones, they will not normally be found in the general population.

In contrast to the high number of false SNPs detected in the EST and UniGene databases, the lack of SNP detection in some genes also constitutes a problem with

these databases. One explanation is that these genes may truly lack polymorphisms, possibly because they possess important physiological functions and are therefore highly conserved [Forsberg *et al.*, 1999]. Alternatively, the problem could lie with the method of EST generation. ESTs are only 300-400 bp in length and, as only the 5' or 3' ends of the cDNA clone are sequenced, the sequence generated does not always cover the entire coding region of a gene of interest. Consistent with a previous study [Irizarry *et al.*, 2000], only a few partial ESTs covered the coding region of the *GSTA4*, *GSTM2*, *GSTM4* and *GSTM5* genes at a particular point. The result is that many polymorphisms may not be detected during database mining. Polymorphisms may also be missed if the transcript of interest is rarely and/or weakly expressed, such as GSTA3-3. Although normalised cDNA libraries have been constructed in order to increase the representation of rare transcripts by reducing the number of highly expressed genes [Soares *et al.*, 1994; Bonaldo *et al.*, 1996], many genes continue to be missed as they are not expressed in the tissues used to create the libraries [Schuler, 1997]. By increasing the overall quality and diversity of the cDNA libraries used to generate ESTs, the quantity of true SNPs in both highly expressed and rare genes is likely to increase, and the number of false positives should correspondingly decrease.

Unlike the publicly available sequence databases, the SNP databases should contain sequence data of a relatively high quality. These databases contain data submitted by various laboratories working specifically on different aspects of SNP discovery. Usually, SNPs must be experimentally validated if they are to be accepted into the databases, for example dbSNP and HGVbase have implemented submission guidelines that ensure high sequence quality [Sherry *et al.*, 2001; Fredman *et al.*, 2002]. EGPSNPs and GeneSNPs both resequence genes of interest generated from the Human Genome Project template using a PDR sample cohort for confirmation, and CGAP divides the polymorphisms predicted using the SNP Finder program into three groups based on different stages of verification. SNP information found in these databases can therefore be expected to be of higher quality than EST data and in turn have a higher likelihood of being real. Therefore, in theory there should be a smaller chance of detecting false positives however, high false positive rates was also a problem encountered when using the SNP databases. This is partly due to the failure of various SNP databases to regularly update and monitor their contents for the presence of SNPs that have subsequently failed further validation tests [Marsh *et al.*, 2002]. In addition, the

presence of paralogues, pseudogenes and duplications of sequences highly similar to the one of interest within the SNP databases can all contribute to the high rate of false positives associated with these databases [Gu *et al.*, 1998; Marth *et al.*, 1999; Marth *et al.*, 2001; Marsh *et al.*, 2002]. One of the main points emphasised by this study is that quite often, genes belonging to multi-gene families, such as the Alpha and Mu class GSTs, are not identified correctly due to high sequence identity with other members of the same gene family. It was found that genes from one gene family are often submitted to the SNP databases as another member of the same gene family, leading to the identification of false polymorphisms. This occurrence has also been observed in other studies based on EST and UniGene searches [Marth *et al.*, 1999; Picoult-Newberg *et al.*, 1999; Irizarry *et al.*, 2000], and one based on the analysis of several SNP databases [Marsh *et al.*, 2002]. Sequence identification problems should become less frequent with the completion of the human genome sequence and its subsequent use as a reference template [Marth *et al.*, 2001].

Implementation of filtering processes during sequence database analysis, and the validity tests in place for submission to the SNP databases, suggests that the sheer number of false positives found in this and other studies cannot be blamed simply on the errors inherent in each individual database. Many of these supposed false positive polymorphisms have high sequence quality, but this and other studies have been unable to further confirm the polymorphisms identified by these means in different populations [Picoult-Newberg *et al.*, 1999; Irizarry *et al.*, 2000]. As discussed in the individual results chapters, there is a possibility that some of these false positives are rare alleles. Rare variants can be detected by database mining, yet these are often excluded and discarded as sequencing errors depending on the degree of filtering enforced. An example is the ranking system found in the SNP Finder program. Potential polymorphisms are given a probability score reflecting the likelihood that the nucleotide in question is truly polymorphic [Buetow *et al.*, 1998]. However, as this score is influenced by the number and quality of sequences found at the polymorphic site, a poorly ranked putative sequence alteration can often be confirmed by electropherographic validation as being a true polymorphism. Depending on the ranking limit set by the user, these rare, but real polymorphisms may be missed. Ultimately, it is up to the researcher to choose between detecting common variants with

a low number of false positives, and risking a higher false positive rate in order to identify rare polymorphisms.

Although rare alleles may be identified and confirmed, they will not necessarily be detected in particular populations. The absence of some of the possibly rare polymorphisms identified in this study in the three ethnic groups screened may be a reflection on the sample size and the choice of population used. Sample sets of 25 individuals for each representative population were used in this study and several other groups have used sample sets ranging from six to 24 individuals in order to confirm the presence of polymorphisms [Forsberg *et al.*, 1998; Buetow *et al.*, 1999; Picoult-Newberg *et al.*, 1999; Lee *et al.*, 2001]. These numbers are sufficient to identify common SNPs [Weiss, 1998], which are found at high frequencies worldwide. These common SNPs are believed to have emerged after speciation approximately 5.5 million years ago [Kumar & Hedges, 1998] but before the divergence of modern human populations, which occurred when humans emerged from Africa approximately one hundred thousand years ago [Cavalli-Sforza *et al.*, 1988; Mountain *et al.*, 1992]. Due to the low rate of genomic change, approximately 85% of the worldwide gene variation is common to all human populations [Barbujani *et al.*, 1997]. The remaining 15% of the global gene variation is composed of rare variants, which tend to occur at a frequency below 5% [Barbujani *et al.*, 1997]. It is believed these developed after human diversification and they are therefore more likely to be restricted to specific ethnic groups [Cargill *et al.*, 1999; Halushka *et al.*, 1999; Risch, 2000]. If these rare alleles are to be detected, larger sample sets representing a diverse number of populations are required. Unfortunately, it is difficult to choose representative populations when trying to confirm SNPs detected through database analysis for the following reasons. Firstly, the ethnicity of samples in the EST and UniGene databases is unknown. Secondly, many polymorphisms submitted to the SNP databases have been identified in PDR sample sets. Although the diversity of this cohort is known, specific details, such as the ethnic origin of particular samples, are withheld in order to prevent misuse of any genetic information that might promote ethnic or racial discrimination [Collins *et al.*, 1998].

The choice of populations used in polymorphism studies must be carefully considered. The four populations used in this study are representative of three major ethnic

populations – the Africans (Bantu and Creole), the South East Asians (Southern Chinese) and the Caucasians. (Australian European). As described above, the modern human population first diversified approximately one hundred thousand years ago when the non-Africans split from the Africans [Cavalli-Sforza *et al.*, 1988; Mountain *et al.*, 1992]. The rest of the global population, including the South East Asians and the Caucasoids (or Europeans) descended from the non-Africans, although the exact order of branch diversification remains unresolved [Cavalli-Sforza *et al.*, 1988; Mountain *et al.*, 1992; Nei & Roychoudhury, 1993]. As mentioned above, two African populations were tested in this study. The Bantu population represents a true African population. In contrast, whilst the Creole population is predominately Bantu African by descent, this group has been subject to mixed Caucasoid (European and Indian) and Asian and Oceania genetic influences [Hewitt *et al.*, 1996].

The majority of the polymorphisms identified in this study were common polymorphisms and minor fluctuations between the allele distributions for the four populations were observed. However, evidence of population specific allele distribution patterns was observed for some of the polymorphisms characterised in this study. The *GSTA3* p.L71 allele was seen in the African populations only, at a frequency of 5% in the Creole group and 15% in the Bantu group. Alleles specific to the African populations are not unusual [Zietkiewicz *et al.*, 1997; Nickerson *et al.*, 1998; Cargill *et al.*, 1999; Halushka *et al.*, 1999], as it is understood that the African populations are genetically more variable than those of non-African descent [Cavalli-Sforza *et al.*, 1994, Zietkiewicz *et al.*, 1997; Clark *et al.*, 1998]. The reduced allele frequency observed in the Creole group may be due to its genetic heterogeneity, and also provides evidence that regional differences within populations must be taken into account in SNP studies. In contrast, the Southern Chinese population showed significantly different *GSTM3* SNP profile patterns to those of the Australian European and African populations, especially that of the Bantu African population. It is possible that this divergence occurred after the South East Asian populations branched from the other non-African populations. Interestingly, these polymorphisms were also found to alter the enzymatic properties of the GSTM3-3 protein. The distributions of these *GSTM3* alleles were similar between the Creole African and Australian European populations, but not the Bantu Africans, which represent an earlier branch of diversification. This similarity between the Creole African and Australian European populations may be due to genetic

similarities between the two populations arising from the fact that the Caucasians are an admixture consisting of 35% ancestral Africans and 65% ancestral Asians [Bowcock *et al.*, 1991], and the Creoles, whilst mainly ancestral African, are also an admixture of Caucasians and Asians [Hewitt *et al.*, 1996]. This may also explain why the *GSTA2* p.T112 and *GSTO2* p.D142 alleles were observed at similar frequencies in the Creole African and Australian European populations, but at frequencies that were lower than those observed in the Bantu African population.

As discussed in the individual results chapters, the polymorphisms identified in this study, particularly those that altered the structure and catalytic activity of the protein, may be implicated in altering the various physiological roles possessed by the Alpha, Mu and Omega class GSTs. The next logical step into characterising the possible physiological effects of these polymorphisms will be to genotype various patient groups presenting symptoms that may be associated with alterations in these different pathways. Based on the findings of previous association studies, it is probable that the polymorphisms identified in this study will not be directly responsible for a particular clinical phenotype, but instead will have an indirect effect through altering susceptibility risks and/or drug responses [Houlston, 1999; Hayes & Strange, 2000; Johns & Houlston, 2000; Strange *et al.*, 2001]. The finding that some GST polymorphisms are associated with the regulation of other GST isoenzymes [Inskip *et al.*, 1995; Coles *et al.*, 2001a] emphasises that factors such as epistatic effects and linkage disequilibrium must be considered during association studies. Strange and colleagues [2001] have suggested that an exploratory patient group be initially genotyped. Confirmatory groups can subsequently be genotyped if the associations achieve significance. In order to do this successfully, the following criteria must be satisfied. Large patient sets from homogenous ethnic and/or geographical backgrounds, with well-matched controls are required. Patients should present with well-defined phenotypes, based on diagnosis following stringent criteria [Houlston, 1999; Hayes & Strange, 2000; Johns & Houlston, 2000; Strange *et al.*, 2000; Lai, 2001].

An important point to consider when assessing the effects these polymorphisms may have on disease aetiology is the variation of allelic distribution between the different ethnic groups. As discussed in Chapter 4, a polymorphism found in 100 individuals from one particular population is not necessarily representative of the rest of that same

population due to the existence of population substructures [Clark *et al.*, 1998]. Although differences in SNP profiles can be expected between similar racial groups located in different countries, regional differences can also exist in one country. Hence differences can also be observed between similar racial groups in the same country [Kawajiri *et al.*, 1990; Park *et al.*, 2000]. Another point is that although similar SNP profiles may be observed in different racial groups, these may be associated with different phenotypes due to differences in linkage disequilibrium, epistasis and other exogenous factors such as environmental and lifestyle effects.

Since this study was completed, the volume of sequence data submitted to the various databases has continued to increase at a great rate. Genes that were poorly represented in the EST and UniGene databases during initial screens in this study are now better represented. For example, the UniGene clusters for the GSTA4 gene now contains 215 EST and mRNA sequences (cf one) and the *GSTM5* gene now has 58 representative sequences (cf 13). The increase in the number of ESTs submitted to EST database, and subsequently collated into clusters in the UniGene database could be attributed to an increase in the number and type of cDNA libraries used to generate ESTs. In 1999, only 200-plus cDNA libraries were available compared to the 440 libraries currently available (6th July 2002). This also represents an increase in the number of individuals represented, therefore increases the probability of finding novel polymorphisms. The various SNP databases are also undergoing constant development, and are now providing general information about the populations in which SNPs have been detected in addition to frequency information. Many, especially dbSNP and GeneSNPs, have newly improved user interfaces. Specifically, the location of any SNP within dbSNP can now be visualised on a colour-coded gene map. Links to ENSEMBL and the UCSC Genome Assembly are also provided for each SNP. One of the newest features however is the haplotype database, which lists the frequency at which haplotypes for particular genes occur in different populations. New features of the GeneSNPs database include a nucleotide map of the gene of interest in which the SNPs are highlighted, and a table containing genotype information for some genes. Others such as the CGAP, EGPSNPs and refseq databases have not undergone significant changes. This continual expansion and development of the EST, UniGene and SNP databases will ensure that these become powerful and reliable tools in the field of SNP detection and, consequently, the application of SNP discovery to a wide range of fields.

References

Abernathy, C. and Morgan, A. (2001). Exposure and health effects. *In*: United Nations Synthesis Report on Arsenic in Drinking Water**:** Chapter 3.

Abramovitz, M., Homma, H., Ishigaki, S., Tansey, F., Cammer, W. and Listowsky, I. (1988). Characterization and localization of glutathione-S-transferases in rat brain and binding of hormones, neurotransmitters, and drugs. *J Neurochem.* **50**(1): 50-7.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F. and et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science.* **252**(5013): 1651-6.

Adler, V., Yin, Z., Fuchs, S. Y., Benezra, M., Rosario, L., Tew, K. D., Pincus, M. R., Sardana, M., Henderson, C. J., Wolf, C. R., Davis, R. J. and Ronai, Z. (1999). Regulation of JNK signaling by GSTp. *Embo J.* **18**(5): 1321-34.

Ahmadi, A., Fredrikson, M., Jerregard, H., Akerback, A., Fall, P. A., Rannug, A., Axelson, O. and Soderkvist, P. (2000). GSTM1 and mEPHX polymorphisms in Parkinson's disease and age of onset. *Biochem Biophys Res Commun.* **269**(3): 676-80.

Aktas, D., Ozen, H., Atsu, N., Tekin, A., Sozen, S. and Tuncbilek, E. (2001). Glutathione S-transferase M1 gene polymorphism in bladder cancer patients. a marker for invasive bladder cancer? *Cancer Genet Cytogenet.* **125**(1): 1-4.

Alarcon, R. A. (1976). Studies on the in vivo formation of acrolein: 3-hydroxy-propylmercapturic acid as an index of cyclophosphamide (NSC-26271) activation. *Cancer Treat Rep.* **60**(4): 327-35.

Alarcon, R. A. and Meienhofer, J. (1971). Formation of the cytotoxic aldehyde acrolein during the in vitro degradation of cyclophosphamide. *Nature (London) New Biology.* **233**: 250-2.

Alexandrie, A. K., Rannug, A., Juronen, E., Tasa, G. and Warholm, M. (2002). Detection and characterization of a novel functional polymorphism in the GSTT1 gene. *Pharmacogenetics.* **12**(8): 613-9.

Ali-Osman, F., Akande, O., Antoun, G., Mao, J. X. and Buolamwini, J. (1997). Molecular cloning, characterization, and expression in Escherichia coli of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. *J Biol Chem.* **272**(15): 10004-12.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17): 3389-402.

Anderer, G., Schrappe, M., Brechlin, A. M., Lauten, M., Muti, P., Welte, K. and Stanulla, M. (2000). Polymorphisms within glutathione S-transferase genes and initial response to glucocorticoids in childhood acute lymphoblastic leukaemia. *Pharmacogenetics.* **10**(8): 715-26.

Anders, M. W., Lash, L., Dekant, W., Elfarra, A. A. and Dohn, D. R. (1988). Biosynthesis and biotransformation of glutathione S-conjugates to toxic metabolites. *Crit Rev Toxicol.* **18**(4): 311-41.

Anderson, W. B., Board, P. G., Gargano, B. and Anders, M. W. (1999). Inactivation of glutathione transferase zeta by dichloroacetic acid and other fluorine-lacking alpha-haloalkanoic acids. *Chem Res Toxicol.* **12**(12): 1144-9.

Ansari, G. A. and Smith, L. L. (1990). Cholesterol epoxides: formation and measurement. *Methods Enzymol.* **186**: 438-43.

Ansari, N. H., Wang, L. and Srivastava, S. K. (1996). Role of lipid aldehydes in cataractogenesis: 4-hydroxynonenal-induced cataract. *Biochem Mol Med.* **58**(1): 25-30.

Anttila, S., Luostarinen, L., Hirvonen, A., Elovaara, E., Karjalainen, A., Nurminen, T., Hayes, J. D., Vainio, H. and Ketterer, B. (1995). Pulmonary expression of glutathione S-transferase M3 in lung cancer patients: association with GSTM1 polymorphism, smoking, and asbestos exposure. *Cancer Res.* **55**(15): 3305-9.

Aposhian, H. V. (1997). Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. *Annu Rev Pharmacol Toxicol.* **37**: 397-419.

Aposhian, H. V., Arroyo, A., Cebrian, M. E., del Razo, L. M., Hurlbut, K. M., Dart, R. C., Gonzalez-Ramirez, D., Kreppel, H., Speisky, H., Smith, A., Gonsebatt, M. E., Ostrosky-Wegman, P. and Aposhian, M. M. (1997). DMPS-arsenic challenge test. I: Increased urinary excretion of monomethylarsonic acid in humans given dimercaptopropane sulfonate. *J Pharmacol Exp Ther.* **282**(1): 192-200.

Aposhian, H. V., Gurzau, E. S., Le, X. C., Gurzau, A., Healy, S. M., Lu, X., Ma, M., Yip, L., Zakharyan, R. A., Maiorino, R. M., Dart, R. C., Tircus, M. G., Gonzalez-Ramirez, D., Morgan, D. L., Avram, D. and Aposhian, M. M. (2000a). Occurrence of monomethylarsonous acid in urine of humans exposed to inorganic arsenic. *Chem Res Toxicol.* **13**(8): 693-7.

Aposhian, H. V., Zheng, B., Aposhian, M. M., Le, X. C., Cebrian, M. E., Cullen, W., Zakharyan, R. A., Ma, M., Dart, R. C., Cheng, Z., Andrewes, P., Yip, L., O'Malley, G. F., Maiorino, R. M., Van Voorhies, W., Healy, S. M. and Titcomb, A. (2000b). DMPS-arsenic challenge test. II. Modulation of arsenic species, including monomethylarsonous acid (MMA(III)), excreted in human urine. *Toxicol Appl Pharmacol.* **165**(1): 74-83.

Armstrong, R. N. (1997). Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem Res Toxicol.* **10**(1): 2-18.

Atsmon, J., Freeman, M. L., Meredith, M. J., Sweetman, B. J. and Roberts, L. J., 2nd (1990a). Conjugation of 9-deoxy-delta 9,delta 12(E)-prostaglandin D2 with intracellular glutathione and enhancement of its antiproliferative activity by glutathione depletion. *Cancer Res.* **50**(6): 1879-85.

Atsmon, J., Sweetman, B. J., Baertschi, S. W., Harris, T. M. and Roberts, L. J., 2nd (1990b). Formation of thiol conjugates of 9-deoxy-delta 9,delta 12(E)-prostaglandin D2 and delta 12(E)-prostaglandin D2. *Biochemistry.* **29**(15): 3760-5.

Awasthi, Y. C., Dao, D. D. and Saneto, R. P. (1980). Interrelationship between anionic and cationic forms of glutathione S-transferases of human liver. *Biochem J.* **191**(1): 1-10.

Babon, J. J., Youil, R. and Cotton, R. G. (1995). Improved strategy for mutation detection--a modification to the enzyme mismatch cleavage method. *Nucleic Acids Res.* **23**(24): 5082-4.

Baez, S., Linderson, Y. and Segura-Aguilar, J. (1995). Superoxide dismutase and catalase enhance autoxidation during one-electron reduction of aminochrome by NADPH-cytochrome P-450 reductase. *Biochem Mol Med.* **54**(1): 12-8.

Baez, S., Segura-Aguilar, J., Widersten, M., Johansson, A. S. and Mannervik, B. (1997). Glutathione transferases catalyse the detoxication of oxidized metabolites (o-quinones) of catecholamines and may serve as an antioxidant system preventing degenerative cellular processes. *Biochem J.* **324 (Pt 1)**: 25-8.

Baker, R. T., Tobias, J. W. and Varshavsky, A. (1992). Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of UBP2 and UBP3, and functional analysis of the UBP gene family. *J Biol Chem.* **267**(32): 23364-75.

Baker, R. T., Smith, S. A., Marano, R., McKee, J. and Board, P. G. (1994). Protein expression using cotranslational fusion and cleavage of ubiquitin. Mutagenesis of the glutathione-binding site of human Pi class glutathione S-transferase. *J Biol Chem.* **269**(41): 25381-6.

Barbujani, G., Magagni, A., Minch, E. and Cavalli-Sforza, L. L. (1997). An apportionment of human DNA diversity. *Proc Natl Acad Sci U S A.* **94**(9): 4516-9.

Barrio, J. R., Secrist, J. A., 3rd and Leonard, N. J. (1972). Fluorescent adenosine and cytidine derivatives. *Biochem Biophys Res Commun.* **46**(2): 597-604.

Barycki, J. J. and Colman, R. F. (1997). Identification of the nonsubstrate steroid binding site of rat liver glutathione S-transferase, isozyme 1-1, by the steroid affinity label, 3beta-(iodoacetoxy)dehydroisoandrosterone. *Arch Biochem Biophys.* **345**(1): 16-31.

Bates, M. N., Smith, A. H. and Hopenhayn-Rich, C. (1992). Arsenic ingestion and internal cancers: a review. *Am J Epidemiol.* **135**(5): 462-76.

Batist, G., Tulpule, A., Sinha, B. K., Katki, A. G., Myers, C. E. and Cowan, K. H. (1986). Overexpression of a novel anionic glutathione transferase in multidrug-resistant human breast cancer cells. *J Biol Chem.* **261**(33): 15544-9.

Baxevanis, A. D. (2002). The Molecular Biology Database Collection: 2002 update. *Nucleic Acids Res.* **30**(1): 1-12.

Beauchamp, R. O., Jr., Andjelkovich, D. A., Kligerman, A. D., Morgan, K. T. and Heck, H. D. (1985). A critical review of the literature on acrolein toxicity. *Crit Rev Toxicol.* **14**(4): 309-80.

Beckett, G. J., Howie, A. F., Hume, R., Matharoo, B., Hiley, C., Jones, P. and Strange, R. C. (1990). Human glutathione S-transferases: radioimmunoassay studies on the expression of alpha-, mu- and pi-class isoenzymes in developing lung and kidney. *Biochim Biophys Acta.* **1036**(3): 176-82.

Beckman, K. B. and Ames, B. N. (1998). The free radical theory of aging matures. *Physiol Rev.* **78**(2): 547-81.

Bell, D. A., Thompson, C. L., Taylor, J., Miller, C. R., Perera, F., Hsieh, L. L. and Lucier, G. W. (1992). Genetic monitoring of human polymorphic cancer susceptibility genes by polymerase chain reaction: application to glutathione transferase mu. *Environ Health Perspect.* **98**: 113-7.

Benhamou, S., Lee, W. J., Alexandrie, A. K., Boffetta, P., Bouchardy, C., Butkiewicz, D., Brockmoller, J., Clapper, M. L., Daly, A., Dolzan, V., Ford, J., Gaspari, L., Haugen, A., Hirvonen, A., Husgafvel-Pursiainen, K., Ingelman-Sundberg, M., Kalina, I., Kihara, M., Kremers, P., Le Marchand, L., London, S. J., Nazar-Stewart, V., Onon-Kihara, M., Rannug, A., Romkes, M., Ryberg, D., Seidegard, J., Shields, P., Strange, R. C., Stucker, I., To-Figueras, J., Brennan, P. and Taioli, E. (2002). Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. *Carcinogenesis.* **23**(8): 1343-50.

Benson, A. M., Talalay, P., Keen, J. H. and Jakoby, W. B. (1977). Relationship between the soluble glutathione-dependent delta 5-3-ketosteroid isomerase and the glutathione S-transferases of the liver. *Proc Natl Acad Sci U S A.* **74**(1): 158-62.

Benson, D. A., Boguski, M., Lipman, D. J. and Ostell, J. (1996). GenBank. *Nucleic Acids Res.* **24**(1): 1-5.

Bera, T. K., Lee, S., Salvatore, G., Lee, B. and Pastan, I. (2001). MRP8, a new member of ABC transporter superfamily, identified by EST database mining and gene prediction program, is highly expressed in breast cancer. *Mol Med.* **7**(8): 509-16.

Bera, T. K., Iavarone, C., Kumar, V., Lee, S., Lee, B. and Pastan, I. (2002a). MRP9, an unusual truncated member of the ABC transporter superfamily, is highly expressed in breast cancer. *Proc Natl Acad Sci U S A.* **99**(10): 6997-7002.

Bera, T. K., Maitra, R., Iavarone, C., Salvatore, G., Kumar, V., Vincent, J. J., Sathyanarayana, B. K., Duray, P., Lee, B. K. and Pastan, I. (2002b). PATE, a gene expressed in prostate cancer, normal prostate, and testis, identified by a functional genomic approach. *Proc Natl Acad Sci U S A.* **99**(5): 3058-63.

Berger, R., Michals, K., Galbraeth, J. and Matalon, R. (1988). Tyrosinemia type Ib caused by maleylacetoacetate isomerase deficiency: A new defect. *Pediatr. Res.* **23**: 328(A).

Berhane, K., Hao, X. Y., Egyhazi, S., Hansson, J., Ringborg, U. and Mannervik, B. (1993). Contribution of glutathione transferase M3-3 to 1,3-bis(2-chloroethyl)-1-nitrosourea resistance in a human non-small cell lung cancer cell line. *Cancer Res.* **53**(18): 4257-61.

Berhane, K., Widersten, M., Engstrom, A., Kozarich, J. W. and Mannervik, B. (1994). Detoxication of base propenals and other alpha, beta-unsaturated aldehyde products of radical reactions and lipid peroxidation by human glutathione transferases. *Proc Natl Acad Sci U S A.* **91**(4): 1480-4.

Berridge, M. J., Bootman, M. D. and Lipp, P. (1998). Calcium--a life and death signal. *Nature.* **395**(6703): 645-8.

Beuckmann, C. T., Fujimori, K., Urade, Y. and Hayaishi, O. (2000). Identification of mu-class glutathione transferases M2-2 and M3-3 as cytosolic prostaglandin E synthases in the human brain. *Neurochem Res.* **25**(5): 733-8.

Beutler, E. (1975). Red blood cells: a manual of biochemical methods. New York, New York: Grune and Stratton.

Bjornestedt, R., Stenberg, G., Widersten, M., Board, P. G., Sinning, I., Jones, T. A. and Mannervik, B. (1995). Functional significance of arginine 15 in the active site of human class alpha glutathione transferase A1-1. *J Mol Biol.* **247**(4): 765-73.

Blackburn, A. C., Woollatt, E., Sutherland, G. R. and Board, P. G. (1998). Characterization and chromosome location of the gene GSTZ1 encoding the human Zeta class glutathione transferase and maleylacetoacetate isomerase. *Cytogenet Cell Genet.* **83**(1-2): 109-14.

Blackburn, A. C., Tzeng, H. F., Anders, M. W. and Board, P. G. (2000). Discovery of a functional polymorphism in human glutathione transferase zeta by expressed sequence tag database analysis. *Pharmacogenetics.* **10**(1): 49-57.

Blackburn, A. C., Coggan, M., Tzeng, H. F., Lantum, H., Polekhina, G., Parker, M. W., Anders, M. W. and Board, P. G. (2001). GSTZ1d: a new allele of glutathione transferase zeta and maleylacetoacetate isomerase. *Pharmacogenetics.* **11**(8): 671-8.

Board, P. G. (1981a). Biochemical genetics of glutathione-S-transferase in man. *Am J Hum Genet.* **33**(1): 36-43.

Board, P. G. (1981b). Gene deletion and partial deficiency of the glutathione S-transferase (ligandin) system in man. *FEBS Lett.* **135**(1): 12-4.

Board, P. G. (1998). Identification of cDNAs encoding two human alpha class glutathione transferases (GSTA3 and GSTA4) and the heterologous expression of GSTA4-4. *Biochem J.* **330 (Pt 2)**: 827-31.

Board, P. G. and Pierce, K. (1987). Expression of human glutathione S-transferase 2 in Escherichia coli. Immunological comparison with the basic glutathione S-transferases isoenzymes from human liver. *Biochem J.* **248**(3): 937-41.

Board, P. G. and Webb, G. C. (1987). Isolation of a cDNA clone and localization of human glutathione S-transferase 2 genes to chromosome band 6p12. *Proc Natl Acad Sci U S A.* **84**(8): 2377-81.

Board, P. G., Webb, G. C. and Coggan, M. (1989). Isolation of a cDNA clone and localization of the human glutathione S-transferase 3 genes to chromosome bands 11q13 and 12q13-14. *Ann Hum Genet.* **53 (Pt 3)**: 205-13.

Board, P., Coggan, M., Johnston, P., Ross, V., Suzuki, T. and Webb, G. (1990). Genetic heterogeneity of the human glutathione transferases: a complex of gene families. *Pharmacol Ther.* **48**(3): 357-69.

Board, P. G., Coggan, M. and Woodcock, D. M. (1992). The human Pi class glutathione transferase sequence at 12q13-q14 is a reverse-transcribed pseudogene. *Genomics.* **14**(2): 470-3.

Board, P. G., Coggan, M., Wilce, M. C. and Parker, M. W. (1995). Evidence for an essential serine residue in the active site of the Theta class glutathione transferases. *Biochem J.* **311 (Pt 1)**: 247-50.

Board, P. G., Baker, R. T., Chelvanayagam, G. and Jermiin, L. S. (1997). Zeta, a novel class of glutathione transferases in a range of species from plants to humans. *Biochem J.* **328 (Pt 3)**: 929-35.

Board, P., Blackburn, A., Jermiin, L. S. and Chelvanayagam, G. (1998). Polymorphism of phase II enzymes: identification of new enzymes and polymorphic variants by database analysis. *Toxicol Lett.* **102-103**: 149-54.

Board, P. G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L. S., Schulte, G. K., Danley, D. E., Hoth, L. R., Griffor, M. C., Kamath, A. V., Rosner, M. H., Chrunyk, B. A., Perregaux, D. E., Gabel, C. A., Geoghegan, K. F. and Pandit, J. (2000). Identification, characterization, and crystal structure of the Omega class glutathione transferases. *J Biol Chem.* **275**(32): 24798-806.

Bogaards, J. J., van Ommen, B. and van Bladeren, P. J. (1993). Interindividual differences in the in vitro conjugation of methylene chloride with glutathione by cytosolic glutathione S-transferase in 22 human liver samples. *Biochem Pharmacol.* **45**(10): 2166-9.

Bogaards, J. J., Verhagen, H., Willems, M. I., van Poppel, G. and van Bladeren, P. J. (1994). Consumption of Brussels sprouts results in elevated alpha-class glutathione S-transferase levels in human blood plasma. *Carcinogenesis.* **15**(5): 1073-5.

Bogaards, J. J., Venekamp, J. C. and van Bladeren, P. J. (1997). Stereoselective conjugation of prostaglandin A2 and prostaglandin J2 with glutathione, catalyzed by the human glutathione S-transferases A1-1, A2-2, M1a-1a, and P1-1. *Chem Res Toxicol.* **10**(3): 310-7.

Boguski, M. S. and Schuler, G. D. (1995). ESTablishing a human transcript map. *Nat Genet.* **10**(4): 369-71.

Boguski, M. S., Lowe, T. M. and Tolstoshev, C. M. (1993). dbEST--database for "expressed sequence tags". *Nat Genet.* **4**(4): 332-3.

Boguski, M., Chakravarti, A., Gibbs, R., Green, E. and Myers, R. M. (1996). The end of the beginning: the race to begin human genome sequencing. *Genome Res.* **6**(9): 771-2.

Bonaldo, M. F., Lennon, G. and Soares, M. B. (1996). Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* **6**(9): 791-806.

Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. and Cavalli-Sforza, L. L. (1991). Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci U S A.* **88**(3): 839-43.

Boyer, T. D., Vessey, D. A., Holcomb, C. and Saley, N. (1984). Studies of the relationship between the catalytic activity and binding of non-substrate ligands by the glutathione S-transferases. *Biochem J.* **217**(1): 179-85.

Bredschneider, M., Klein, K., Murdter, T. E., Marx, C., Eichelbaum, M., Nussler, A. K., Neuhaus, P., Zanger, U. M. and Schwab, M. (2002). Genetic polymorphisms of glutathione S-transferase A1, the major glutathione S-transferase in human liver: consequences for enzyme expression and busulfan conjugation. *Clin Pharmacol Ther.* **71**(6): 479-87.

Brockmoller, J., Kerb, R., Drakoulis, N., Staffeldt, B. and Roots, I. (1994). Glutathione S-transferase M1 and its variants A and B as host factors of bladder cancer susceptibility: a case-control study. *Cancer Res.* **54**(15): 4103-11.

Brockmoller, J., Cascorbi, I., Kerb, R., Sachse, C. and Roots, I. (1998). Polymorphisms in xenobiotic conjugation and disease predisposition. *Toxicol Lett.* **102-103**: 173-83.

Brookes, A. J. (1999). The essence of SNPs. *Gene.* **234**(2): 177-86.

Brookes, A. J. (2002). 4th International Meeting on Single Nucleotide Polymorphism and Complex Genome AnalysisVarious Uses for DNA Variations. *Eur J Hum Genet.* **10**(2): 153-5.

Brookes, A. J., Lehvaslaiho, H., Siegfried, M., Boehm, J. G., Yuan, Y. P., Sarkar, C. M., Bork, P. and Ortigao, F. (2000). HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res.* **28**(1): 356-60.

Brophy, P. M., Southan, C. and Barrett, J. (1989). Glutathione transferases in the tapeworm Moniezia expansa. *Biochem J.* **262**(3): 939-46.

Bruns, C. M., Hubatsch, I., Ridderstrom, M., Mannervik, B. and Tainer, J. A. (1999). Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J Mol Biol.* **288**(3): 427-39.

Buetler, T. M. and Eaton, D. L. (1992). Glutathione-S-tranferases - amino acid sequence comparison, classification and phylogenetic relationship. *Journal of Environmental Science and Health, Part C.* **Environmental Carcinogenesis and Ecotoxicology Reviews, 10**(2): 181-203.

Buetow, K. H., Edmonson, M. N. and Cassidy, A. B. (1999). Reliable identification of large numbers of candidate SNPs from public EST data. *Nat Genet.* **21**(3): 323-5.

Bull, R. J., Sanchez, I. M., Nelson, M. A., Larson, J. L. and Lansing, A. J. (1990). Liver tumor induction in B6C3F1 mice by dichloroacetate and trichloroacetate. *Toxicology.* **63**(3): 341-59.

Burek, J. D., Nitschke, K. D., Bell, T. J., Wackerle, D. L., Childs, R. C., Beyer, J. E., Dittenber, D. A., Rampy, L. W. and McKenna, M. J. (1984). Methylene chloride: a two-year inhalation toxicity and oncogenicity study in rats and hamsters. *Fundam Appl Toxicol.* **4**(1): 30-47.

Burgess, J. R., Chow, N. W., Reddy, C. C. and Tu, C. P. (1989). Amino acid substitutions in the human glutathione S-transferases confer different specificities in the prostaglandin endoperoxide conversion pathway. *Biochem Biophys Res Commun.* **158**(2): 497-502.

Caccuri, A. M., Antonini, G., Board, P. G., Flanagan, J., Parker, M. W., Paolesse, R., Turella, P., Federici, G., Lo Bello, M. and Ricci, G. (2001). Human glutathione transferase T2-2 discloses some evolutionary strategies for optimization of substrate binding to the active site of glutathione transferases. *J Biol Chem.* **276**(8): 5427-31.

Cammer, W., Tansey, F., Abramovitz, M., Ishigaki, S. and Listowsky, I. (1989). Differential localization of glutathione-S-transferase Yp and Yb subunits in oligodendrocytes and astrocytes of rat brain. *J Neurochem.* **52**(3): 876-83.

Campbell, E., Takahashi, Y., Abramovitz, M., Peretz, M. and Listowsky, I. (1990). A distinct human testis and brain mu-class glutathione S-transferase. Molecular cloning and characterization of a form present even in individuals lacking hepatic type mu isoenzymes. *J Biol Chem.* **265**(16): 9188-93.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q. and Lander, E. S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* **22**(3): 231-8.

Castro, V. M., Soderstrom, M., Carlberg, I., Widersten, M., Platz, A. and Mannervik, B. (1990). Differences among human tumor cell lines in the expression of glutathione transferases and other glutathione-linked enzymes. *Carcinogenesis.* **11**(9): 1569-76.

Cavalli-Sforza, L. L. (1998). The DNA revolution in population genetics. *Trends Genet.* **14**(2): 60-5.

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. and Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci U S A.* **85**(16): 6002-6.

Cavalli-Sforza, L. L., Menozzi, P. and Piazza, A. (1994). The History and Geography of Human Genes. Princeton, New Jersey, Princeton University Press.

Cebrian, M. E., Albores, A., Aguilar, M. and Blakely, E. (1983). Chronic arsenic poisoning in the north of Mexico. *Hum Toxicol.* **2**(1): 121-33.

Chang, J. C. and Kan, Y. W. (1979). beta 0 thalassemia, a nonsense mutation in man. *Proc Natl Acad Sci U S A.* **76**(6): 2886-9.

Chen, L. Z. and Board, P. G. (1987). Hgi AI restriction fragment length polymorphism at the human glutathione S-transferase 2 locus. *Nucleic Acids Res.* **15**(15): 6306.

Chen, H. and Juchau, M. R. (1998). Recombinant human glutathione S-transferases catalyse enzymic isomerization of 13-cis-retinoic acid to all-trans-retinoic acid in vitro. *Biochem J.* **336 (Pt 1)**: 223-6.

Chen, C. J., Chuang, Y. C., Lin, T. M. and Wu, H. Y. (1985). Malignant neoplasms among residents of a blackfoot disease-endemic area in Taiwan: high-arsenic artesian well water and cancers. *Cancer Res.* **45**(11 Pt 2): 5895-9.

Chen, C. L., Liu, Q., Pui, C. H., Rivera, G. K., Sandlund, J. T., Ribeiro, R., Evans, W. E. and Relling, M. V. (1997). Higher frequency of glutathione S-transferase deletions in black children with acute lymphoblastic leukemia. *Blood.* **89**(5): 1701-7.

Cho, S. G., Lee, Y. H., Park, H. S., Ryoo, K., Kang, K. W., Park, J., Eom, S. J., Kim, M. J., Chang, T. S., Choi, S. Y., Shim, J., Kim, Y., Dong, M. S., Lee, M. J., Kim, S. G., Ichijo, H. and Choi, E. J. (2001). Glutathione S-transferase mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. *J Biol Chem.* **276**(16): 12749-55.

Chow, N. W., Whang-Peng, J., Kao-Shan, C. S., Tam, M. F., Lai, H. C. and Tu, C. P. (1988). Human glutathione S-transferases. The Ha multigene family encodes products of different but overlapping substrate specificities. *J Biol Chem.* **263**(26): 12797-800.

Ciaccio, P. J., Tew, K. D. and LaCreta, F. P. (1991). Enzymatic conjugation of chlorambucil with glutathione by human glutathione S-transferases and inhibition by ethacrynic acid. *Biochem Pharmacol.* **42**(7): 1504-7.

Clark, A. G., Smith, J. N. and Speir, T. W. (1973). Cross specificity in some vertebrate and insect glutathione-transferases with methyl parathion (dimethyl p-nitrophenyl phosphorothionate), 1-chloro-2,4-dinitro-benzene and s-crotonyl-N-acetylcysteamine as substrates. *Biochem J.* **135**(3): 385-92.

Clark, A. G., Weiss, K. M., Nickerson, D. A., Taylor, S. L., Buchanan, A., Stengard, J., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E. and Sing, C. F. (1998). Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am J Hum Genet.* **63**(2): 595-612.

Coan, C., Ji, J. Y., Hideg, K. and Mehlhorn, R. J. (1992). Protein sulfhydryls are protected from irreversible oxidation by conversion to mixed disulfides. *Arch Biochem Biophys.* **295**(2): 369-78.

Coceani, F. and Akarsu, E. S. (1998). Prostaglandin E2 in the pathogenesis of fever. An update. *Ann N Y Acad Sci.* **856**: 76-82.

Coggan, M., Whitbread, L., Whittington, A. and Board, P. (1998). Structure and organization of the human theta-class glutathione S-transferase and D-dopachrome tautomerase gene complex. *Biochem J.* **334 (Pt 3)**: 617-23.

Cole, S. P., Bhardwaj, G., Gerlach, J. H., Mackie, J. E., Grant, C. E., Almquist, K. C., Stewart, A. J., Kurz, E. U., Duncan, A. M. and Deeley, R. G. (1992). Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science.* **258**(5088): 1650-4.

Coles, B. and Ketterer, B. (1990). The role of glutathione and glutathione transferases in chemical carcinogenesis. *Crit Rev Biochem Mol Biol.* **25**(1): 47-70.

Coles, B. F., Anderson, K. E., Doerge, D. R., Churchwell, M. I., Lang, N. P. and Kadlubar, F. F. (2000). Quantitative analysis of interindividual variation of glutathione S-transferase expression in human pancreas and the ambiguity of correlating genotype with phenotype. *Cancer Res.* **60**(3): 573-9.

Coles, B. F., Morel, F., Rauch, C., Huber, W. W., Yang, M., Teitel, C. H., Green, B., Lang, N. P. and Kadlubar, F. F. (2001a). Effect of polymorphism in the human glutathione S-transferase A1 promoter on hepatic GSTA1 and GSTA2 expression. *Pharmacogenetics.* **11**(8): 663-9.

Coles, B., Nowell, S. A., MacLeod, S. L., Sweeney, C., Lang, N. P. and Kadlubar, F. F. (2001b). The role of human glutathione S-transferases (hGSTs) in the detoxification of the food-derived carcinogen metabolite N-acetoxy-PhIP, and the effect of a polymorphism in hGSTA1 on colorectal cancer risk. *Mutat Res.* **482**(1-2): 3-10.

Coles, B. F., Chen, G., Kadlubar, F. F. and Radominska-Pandya, A. (2002). Interindividual variation and organ-specific patterns of glutathione S-transferase alpha, mu, and pi expression in gastrointestinal tract mucosa of normal individuals. *Arch Biochem Biophys.* **403**(2): 270-6.

Collins, F. S. (1999). Shattuck lecture--medical and societal consequences of the Human Genome Project. *N Engl J Med.* **341**(1): 28-37.

Collins, F. and Galas, D. (1993). A new five-year plan for the U.S. Human Genome Project. *Science.* **262**(5130): 43-6.

Collins, F. S., Brooks, L. D. and Chakravarti, A. (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**(12): 1229-31.

Comstock, K. E., Johnson, K. J., Rifenbery, D. and Henner, W. D. (1993). Isolation and analysis of the gene and cDNA for a human Mu class glutathione S-transferase, GSTM4. *J Biol Chem.* **268**(23): 16958-65.

Comstock, K. E., Widersten, M., Hao, X. Y., Henner, W. D. and Mannervik, B. (1994). A comparison of the enzymatic and physicochemical properties of human glutathione transferase M4-4 and three other human Mu class enzymes. *Arch Biochem Biophys.* **311**(2): 487-95.

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. and Schmidtke, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet.* **69**(3): 201-5.

Cornett, R., James, M. O., Henderson, G. N., Cheung, J., Shroads, A. L. and Stacpoole, P. W. (1999). Inhibition of glutathione S-transferase zeta and tyrosine metabolism by dichloroacetate: a potential unifying mechanism for its altered biotransformation and toxicity. *Biochem Biophys Res Commun.* **262**(3): 752-6.

Cortessis, V., Siegmund, K., Chen, Q., Zhou, N., Diep, A., Frankl, H., Lee, E., Zhu, Q. S., Haile, R. and Levy, D. (2001). A case-control study of microsomal epoxide hydrolase, smoking, meat consumption, glutathione S-transferase M3, and risk of colorectal adenomas. *Cancer Res.* **61**(6): 2381-5.

Cossar, D., Bell, J., Strange, R., Jones, M., Sandison, A. and Hume, R. (1990). The alpha and pi isoenzymes of glutathione S-transferase in human fetal lung: in utero ontogeny compared with differentiation in lung organ culture. *Biochim Biophys Acta.* **1037**(2): 221-6.

Cotton, R. G., Rodrigues, N. R. and Campbell, R. D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc Natl Acad Sci U S A.* **85**(12): 4397-401.

Cotton, R. G., McKusick, V. and Scriver, C. R. (1998). The HUGO Mutation Database Initiative. *Science.* **279**(5347): 10-1.

Cowell, I. G., Dixon, K. H., Pemble, S. E., Ketterer, B. and Taylor, J. B. (1988). The structure of the human glutathione S-transferase pi gene. *Biochem J.* **255**(1): 79-83.

Crofts, F., Cosma, G. N., Currie, D., Taioli, E., Toniolo, P. and Garte, S. J. (1993). A novel CYP1A1 gene polymorphism in African-Americans. *Carcinogenesis.* **14**(9): 1729-31.

Dagnino-Subiabre, A., Cassels, B. K., Baez, S., Johansson, A. S., Mannervik, B. and Segura-Aguilar, J. (2000). Glutathione transferase M2-2 catalyzes conjugation of dopamine and dopa o-quinones. *Biochem Biophys Res Commun.* **274**(1): 32-6.

Dainelli, B., Paludi, D., Dragani, B., Cocco, R., Principe, D. R., Petrucci, M., Mucilli, F., Faraone, A. and Aceto, A. (2002). A novel glutathione transferase from Haemophilus influenzae which has high affinity towards antibiotics. *Int J Biochem Cell Biol.* **34**(8): 916-20.

Daly, A. K., Thomas, D. J., Cooper, J., Pearson, W. R., Neal, D. E. and Idle, J. R. (1993). Homozygous deletion of gene for glutathione S-transferase M1 in bladder cancer. *Bmj.* **307**(6902): 481-2.

Danielson, U. H. and Mannervik, B. (1985). Kinetic independence of the subunits of cytosolic glutathione transferase from the rat. *Biochem J.* **231**(2): 263-7.

Danielson, U. H., Esterbauer, H. and Mannervik, B. (1987). Structure-activity relationships of 4-hydroxyalkenals in the conjugation catalysed by mammalian glutathione transferases. *Biochem J.* **247**(3): 707-13.

Davies, S. M., Robison, L. L., Buckley, J. D., Radloff, G. A., Ross, J. A. and Perentesis, J. P. (2000). Glutathione S-transferase polymorphisms in children with myeloid leukemia: a Children's Cancer Group study. *Cancer Epidemiol Biomarkers Prev.* **9**(6): 563-6.

de Bruin, W. C., Wagenmans, M. J., Board, P. G. and Peters, W. H. (1999). Expression of glutathione S-transferase theta class isoenzymes in human colorectal and gastric cancers. *Carcinogenesis.* **20**(8): 1453-7.

DeAngelo, A. B., Daniel, F. B., Most, B. M. and Olson, G. R. (1996). The carcinogenicity of dichloroacetic acid in the male Fischer 344 rat. *Toxicology.* **114**(3): 207-21.

DeJong, J. L., Chang, C. M., Whang-Peng, J., Knutsen, T. and Tu, C. P. (1988). The human liver glutathione S-transferase gene superfamily: expression and chromosome mapping of an Hb subunit cDNA. *Nucleic Acids Res.* **16**(17): 8541-54.

DeJong, J. L., Mohandas, T. and Tu, C. P. (1991). The human Hb (mu) class glutathione S-transferases are encoded by a dispersed gene family. *Biochem Biophys Res Commun.* **180**(1): 15-22.

Del Corso, A., Cappiello, M. and Mura, U. (1994). Thiol dependent oxidation of enzymes: the last chance against oxidative stress. *Int J Biochem.* **26**(6): 745-50.

den Dunnen, J. T. and Antonarakis, S. E. (2001). Nomenclature for the description of human sequence variations. *Hum Genet.* **109**(1): 121-4.

Desmots, F., Rauch, C., Henry, C., Guillouzo, A. and Morel, F. (1998). Genomic organization, 5'-flanking region and chromosomal localization of the human glutathione transferase A4 gene. *Biochem J.* **336 (Pt 2)**: 437-42.

Desmots, F., Rissel, M., Loyer, P., Turlin, B. and Guillouzo, A. (2001). Immunohistological analysis of glutathione transferase A4 distribution in several human tissues using a specific polyclonal antibody. *J Histochem Cytochem.* **49**(12): 1573-80.

Di Ilio, C., Aceto, A., Bucciarelli, T., Angelucci, S., Felaco, M., Grilli, A. and Federici, G. (1990). Glutathione transferase isoenzymes from human prostate. *Biochem J.* **271**(2): 481-5.

Dirr, H., Reinemer, P. and Huber, R. (1994). X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur J Biochem.* **220**(3): 645-61.

Dirven, H. A., van Ommen, B. and van Bladeren, P. J. (1994). Involvement of human glutathione S-transferase isoenzymes in the conjugation of cyclophosphamide metabolites with glutathione. *Cancer Res.* **54**(23): 6215-20.

Doll, R. and Peto, R. (1981). The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst.* **66**(6): 1191-308.

Dorion, S., Lambert, H. and Landry, J. (2002). Activation of the p38 signaling pathway by heat shock involves the dissociation of glutathione-S-transferase mu from Ask1. *J Biol Chem.* **277**(34): 30792-7.

Dulhunty, A., Haarmann, C., Green, D. and Hart, J. (2000). How many cysteine residues regulate ryanodine receptor channel activity? *Antioxid Redox Signal.* **2**(1): 27-34.

Dulhunty, A., Gage, P., Curtis, S., Chelvanayagam, G. and Board, P. (2001). The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator. *J Biol Chem.* **276**(5): 3319-23.

Eckl, P. M., Ortner, A. and Esterbauer, H. (1993). Genotoxic properties of 4-hydroxyalkenals and analogous aldehydes. *Mutat Res.* **290**(2): 183-92.

Edwards, A., Gibbs, R. A., Nguyen, P. N., Ansorge, W. and Caskey, C. T. (1989). Automated DNA sequencing methods for detection and analysis of mutations: applications to the Lesch-Nyhan syndrome. *Trans Assoc Am Physicians.* **102**: 185-94.

Edwards, A., Civitello, A., Hammond, H. A. and Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet.* **49**(4): 746-56.

Egyhazi, S., Edgren, M. R., Hansson, J., Krockel, D., Mannervik, B. and Ringborg, U. (1997). Role of O6-methylguanine-DNA methyltransferase, glutathione transferase M3-3 and glutathione in resistance to carmustine in a human non-small cell lung cancer cell line. *Eur J Cancer.* **33**(3): 447-52.

Emahazion, T., Jobs, M., Howell, W. M., Siegfried, M., Wyoni, P. I., Prince, J. A. and Brookes, A. J. (1999). Identification of 167 polymorphisms in 88 genes from candidate neurodegeneration pathways. *Gene.* **238**(2): 315-24.

Engel, L. S., Taioli, E., Pfeiffer, R., Garcia-Closas, M., Marcus, P. M., Lan, Q., Boffetta, P., Vineis, P., Autrup, H., Bell, D. A., Branch, R. A., Brockmoller, J., Daly, A. K., Heckbert, S. R., Kalina, I., Kang, D., Katoh, T., Lafuente, A., Lin, H. J., Romkes, M., Taylor, J. A. and Rothman, N. (2002). Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review. *Am J Epidemiol.* **156**(2): 95-109.

Engels, W. R. (1993). Contributing software to the internet: the Amplify program. *Trends Biochem Sci.* **18**(11): 448-50.

Esterbauer, H., Schaur, R. J. and Zollner, H. (1991). Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radic Biol Med.* **11**(1): 81-128.

Evans, W. E. and Relling, M. V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science.* **286**(5439): 487-91.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**(3): 186-94.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**(3): 175-85.

Faulder, C. G., Hirrell, P. A., Hume, R. and Strange, R. C. (1987). Studies of the development of basic, neutral and acidic isoenzymes of glutathione S-transferase in human liver, adrenal, kidney and spleen. *Biochem J.* **241**(1): 221-8.

Felton, J. S., Knize, M. G., Dolbeare, F. A. and Wu, R. (1994). Mutagenic activity of heterocyclic amines in cooked foods. *Environ Health Perspect.* **102 Suppl 6**: 201-4.

Fernandez-Canon, J. M. and Penalva, M. A. (1998). Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. *J Biol Chem.* **273**(1): 329-37.

Fernandez-Canon, J. M., Hejna, J., Reifsteck, C., Olson, S. and Grompe, M. (1999). Gene structure, chromosomal location, and expression pattern of maleylacetoacetate isomerase. *Genomics.* **58**(3): 263-9.

Flamigni, F., Marmiroli, S., Caldarera, C. M. and Guarnieri, C. (1989). Involvement of thiol transferase- and thioredoxin-dependent systems in the protection of 'essential' thiol groups of ornithine decarboxylase. *Biochem J.* **259**(1): 111-5.

Forman, B. M., Tontonoz, P., Chen, J., Brun, R. P., Spiegelman, B. M. and Evans, R. M. (1995). 15-Deoxy-delta 12, 14-prostaglandin J2 is a ligand for the adipocyte determination factor PPAR gamma. *Cell.* **83**(5): 803-12.

Forsberg, L., de Faire, U. and Morgenstern, R. (1999). Low yield of polymorphisms from EST blast searching: analysis of genes related to oxidative stress and verification of the P197L polymorphism in GPX1. *Hum Mutat.* **13**(4): 294-300.

Fournier, D., Bride, J. M., Poirie, M., Berge, J. B. and Plapp, F. W., Jr. (1992). Insect glutathione S-transferases. Biochemical characteristics of the major forms from houseflies susceptible and resistant to insecticides. *J Biol Chem.* **267**(3): 1840-5.

Fredman, D., Siegfried, M., Yuan, Y. P., Bork, P., Lehvaslaiho, H. and Brookes, A. J. (2002). HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. *Nucleic Acids Res.* **30**(1): 387-91.

Frezal, J. (1998). Genatlas database, genes and development defects. *C R Acad Sci III.* **321**(10): 805-17.

Gardner, J. L. and Gallagher, E. P. (2001). Development of a peptide antibody specific to human glutathione S-transferase alpha 4-4 (hGSTA4-4) reveals preferential localization in human liver mitochondria. *Arch Biochem Biophys.* **390**(1): 19-27.

Garg, K., Green, P. and Nickerson, D. A. (1999). Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Res.* **9**(11): 1087-92.

Garte, S. (1998). The role of ethnicity in cancer susceptibility gene polymorphisms: the example of CYP1A1. *Carcinogenesis.* **19**(8): 1329-32.

Gebhardt, R. (1992). Metabolic zonation of the liver: regulation and implications for liver function. *Pharmacol Ther.* **53**(3): 275-354.

Gertig, D. M., Stampfer, M., Haiman, C., Hennekens, C. H., Kelsey, K. and Hunter, D. J. (1998). Glutathione S-transferase GSTM1 and GSTT1 polymorphisms and colorectal cancer risk: a prospective study. *Cancer Epidemiol Biomarkers Prev.* **7**(11): 1001-5.

Giloni, L., Takeshita, M., Johnson, F., Iden, C. and Grollman, A. P. (1981). Bleomycin-induced strand-scission of DNA. Mechanism of deoxyribose cleavage. *J Biol Chem.* **256**(16): 8608-15.

Gordon, D., Abajian, C. and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* **8**(3): 195-202.

Grollman, A. P., Takeshita, M., Pillai, K. M. and Johnson, F. (1985). Origin and cytotoxic properties of base propenals derived from DNA. *Cancer Res.* **45**(3): 1127-31.

Group, The Huntington's Disease Collaborative Research. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell.* **72**(6): 971-83.

Group, The International SNP Map Working. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* **409**(6822): 928-33.

Gu, Z., Hillier, L. and Kwok, P. Y. (1998). Single nucleotide polymorphism hunting in cyberspace. *Hum Mutat.* **12**(4): 221-5.

Guengerich, F. P. and Persmark, M. (1994). Mechanism of formation of ethenoguanine adducts from 2-haloacetaldehydes: 13C-labeling patterns with 2-bromoacetaldehyde. *Chem Res Toxicol.* **7**(2): 205-8.

Guex, N. and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* **18**(15): 2714-23.

Guha Mazumder, D. N., Chakraborty, A. K., Ghose, A., Gupta, J. D., Chakraborty, D. P., Dey, S. B. and Chattopadhyay, N. (1988). Chronic arsenic toxicity from drinking tubewell water in rural West Bengal. *Bull World Health Organ.* **66**(4): 499-506.

Guillaudeux, T., Janer, M., Wong, G. K., Spies, T. and Geraghty, D. E. (1998). The complete genomic sequence of 424,015 bp at the centromeric end of the HLA class I region: gene content and polymorphism. *Proc Natl Acad Sci U S A.* **95**(16): 9494-9.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y. and et al. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature.* **306**(5940): 234-8.

Gustafsson, A., Etahadieh, M., Jemth, P. and Mannervik, B. (1999). The C-terminal region of human glutathione transferase A1-1 affects the rate of glutathione binding and the ionization of the active-site Tyr9. *Biochemistry.* **38**(49): 16268-75.

Gustafsson, A., Nilsson, L. O. and Mannervik, B. (2002). Hybridization of alpha class subunits generating a functional glutathione transferase A1-4 heterodimer. *J Mol Biol.* **316**(2): 395-406.

Gut, I. G. (2001). Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat.* **17**(6): 475-92.

Guthenberg, C., Warholm, M., Rane, A. and Mannervik, B. (1986). Two distinct forms of glutathione transferase from human foetal liver. Purification and comparison with isoenzymes isolated from adult liver and placenta. *Biochem J.* **235**(3): 741-5.

Habig, W. H. and Jakoby, W. B. (1981). Assays for differentiation of glutathione S-transferases. *Methods Enzymol.* **77**: 398-405.

Hall, A. G., Autzen, P., Cattan, A. R., Malcolm, A. J., Cole, M., Kernahan, J. and Reid, M. M. (1994). Expression of mu class glutathione S-transferase correlates with event-free survival in childhood acute lymphoblastic leukemia. *Cancer Res.* **54**(20): 5251-4.

Hallier, E., Langhof, T., Dannappel, D., Leutbecher, M., Schroder, K., Goergens, H. W., Muller, A. and Bolt, H. M. (1993). Polymorphism of glutathione conjugation of methyl bromide, ethylene oxide and dichloromethane in human blood: influence on the induction of sister chromatid exchanges (SCE) in lymphocytes. *Arch Toxicol.* **67**(3): 173-8.

Hallier, E., Schroder, K. R., Asmuth, K., Dommermuth, A., Aust, B. and Goergens, H. W. (1994). Metabolism of dichloromethane (methylene chloride) to formaldehyde in human erythrocytes: influence of polymorphism of glutathione transferase theta (GST T1-1). *Arch Toxicol.* **68**(7): 423-7.

Halliwell, B., Wasil, M. and Grootveld, M. (1987). Biologically significant scavenging of the myeloperoxidase-derived oxidant hypochlorous acid by ascorbic acid. Implications for antioxidant protection in the inflamed rheumatoid joint. *FEBS Lett.* **213**(1): 15-7.

Halushka, M. K., Fan, J. B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. and Chakravarti, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet.* **22**(3): 239-47.

Hamosh, A., Scott, A. F., Amberger, J., Valle, D. and McKusick, V. A. (2000). Online Mendelian Inheritance in Man (OMIM). *Hum Mutat.* **15**(1): 57-61.

Hanson, S. R., Chen, A. A., Smith, J. B. and Lou, M. F. (1999). Thiolation of the gammaB-crystallins in intact bovine lens exposed to hydrogen peroxide. *J Biol Chem.* **274**(8): 4735-42.

Hao, X. Y., Bergh, J., Brodin, O., Hellman, U. and Mannervik, B. (1994). Acquired resistance to cisplatin and doxorubicin in a small cell lung cancer cell line is correlated to elevated expression of glutathione-linked detoxification enzymes. *Carcinogenesis.* **15**(6): 1167-73.

Harada, S., Fujii, C., Hayashi, A. and Ohkoshi, N. (2001). An association between idiopathic Parkinson's disease and polymorphisms of phase II detoxification enzymes: glutathione S-transferase M1 and quinone oxidoreductase 1 and 2. *Biochem Biophys Res Commun.* **288**(4): 887-92.

Hariharan, N., Kelley, D. E. and Perry, R. P. (1991). Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *Proc Natl Acad Sci U S A.* **88**(21): 9799-803.

Harrington, J. J. and Lieber, M. R. (1994). The characterization of a mammalian DNA structure-specific endonuclease. *Embo J.* **13**(5): 1235-46.

Harris, J. M., Meyer, D. J., Coles, B. and Ketterer, B. (1991). A novel glutathione transferase (13-13) isolated from the matrix of rat liver mitochondria having structural similarity to class theta enzymes. *Biochem J.* **278 (Pt 1)**: 137-41.

Hayaishi, O. (1988). Sleep-wake regulation by prostaglandins D2 and E2. *J Biol Chem.* **263**(29): 14593-6.

Hayashi, S., Watanabe, J. and Kawajiri, K. (1992). High susceptibility to lung cancer analyzed in terms of combined genotypes of P450IA1 and Mu-class glutathione S-transferase genes. *Jpn J Cancer Res.* **83**(8): 866-70.

Hayes, J. D. and McLellan, L. I. (1999). Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. *Free Radic Res.* **31**(4): 273-300.

Hayes, J. D. and Pulford, D. J. (1995). The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Biol.* **30**(6): 445-600.

Hayes, J. D. and Strange, R. C. (2000). Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology.* **61**(3): 154-66.

Hayes, J. D. and Wolf, C. R. (1990). Molecular mechanisms of drug resistance. *Biochem J.* **272**(2): 281-95.

Hayes, J. D., Strange, R. C. and Percy-Robb, I. W. (1980). Cholic acid binding by glutathione S-transferases from rat liver cytosol. *Biochem J.* **185**(1): 83-7.

Hayes, J. D., Kerr, L. A. and Cronshaw, A. D. (1989). Evidence that glutathione S-transferases B1B1 and B2B2 are the products of separate genes and that their expression in human liver is subject to inter-individual variation. Molecular relationships between the B1 and B2 subunits and other Alpha class glutathione S-transferases. *Biochem J.* **264**(2): 437-45.

Hewitt, R., Krause, A., Goldman, A., Campbell, G. and Jenkins, T. (1996). Beta-globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking Negroids. *Am J Hum Genet.* **58**(6): 1303-8.

Hietanen, E., Husgafvel-Pursiainen, K. and Vainio, H. (1997). Interaction between dose and susceptibility to environmental cancer: a short review. *Environ Health Perspect.* **105 Suppl 4**: 749-54.

Hiley, C., Fryer, A., Bell, J., Hume, R. and Strange, R. C. (1988). The human glutathione S-transferases. Immunohistochemical studies of the developmental expression of Alpha- and Pi-class isoenzymes in liver. *Biochem J.* **254**(1): 255-9.

Hillier, L. D., Lennon, G., Becker, M., Bonaldo, M. F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Marra, M. and et al. (1996). Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**(9): 807-28.

Hiratsuka, A., Sebata, N., Kawashima, K., Okuda, H., Ogura, K., Watabe, T., Satoh, K., Hatayama, I., Tsuchida, S., Ishikawa, T. and et al. (1990). A new class of rat glutathione S-transferase Yrs-Yrs inactivating reactive sulfate esters as metabolites of carcinogenic arylmethanols. *J Biol Chem.* **265**(20): 11973-81.

Hirvonen, A. (1999). Polymorphisms of xenobiotic-metabolizing enzymes and susceptibility to cancer. *Environ Health Perspect.* **107 Suppl 1**: 37-47.

Hirvonen, A., Husgafvel-Pursiainen, K., Karjalainen, A., Anttila, S. and Vainio, H. (1992). Point-mutational MspI and Ile-Val polymorphisms closely linked in the CYP1A1 gene: lack of association with susceptibility to lung cancer in a Finnish study population. *Cancer Epidemiol Biomarkers Prev.* **1**(6): 485-9.

Hirvonen, A., Husgafvel-Pursiainen, K., Anttila, S. and Vainio, H. (1993). The GSTM1 null genotype as a potential risk modifier for squamous cell carcinoma of the lung. *Carcinogenesis.* **14**(7): 1479-81.

Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A.* **88**(16): 7276-80.

Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. *J Biol Chem.* **264**(24): 13963-6.

Holmgren, A. (2000). Antioxidant function of thioredoxin and glutaredoxin systems. *Antioxid Redox Signal.* **2**(4): 811-20.

Holmgren, A. and Aslund, F. (1995). Glutaredoxin. *Methods Enzymol.* **252**: 283-92.

Homma, H. and Listowsky, I. (1985). Identification of Yb-glutathione-S-transferase as a major rat liver protein labeled with dexamethasone 21-methanesulfonate. *Proc Natl Acad Sci U S A.* **82**(21): 7165-9.

Hopper, E., Belinsky, M. G., Zeng, H., Tosolini, A., Testa, J. R. and Kruh, G. D. (2001). Analysis of the structure and expression pattern of MRP7 (ABCC10), a new member of the MRP subfamily. *Cancer Lett.* **162**(2): 181-91.

Ingram, V. M. (1957). Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature*. 180: 326-8.

Horton, R., Niblett, D., Milne, S., Palmer, S., Tubby, B., Trowsdale, J. and Beck, S. (1998). Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. *J Mol Biol.* **282**(1): 71-97.

Houlgatte, R., Mariage-Samson, R., Duprat, S., Tessier, A., Bentolila, S., Lamy, B. and Auffray, C. (1995). The Genexpress Index: a resource for gene discovery and the genic map of the human genome. *Genome Res.* **5**(3): 272-304.

Houlston, R. S. (1999). Glutathione S-transferase M1 status and lung cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* **8**(8): 675-82.

Hu, X., Ji, X., Srivastava, S. K., Xia, H., Awasthi, S., Nanduri, B., Awasthi, Y. C., Zimniak, P. and Singh, S. V. (1997a). Mechanism of differential catalytic efficiency of two polymorphic forms of human glutathione S-transferase P1-1 in the glutathione conjugation of carcinogenic diol epoxide of chrysene. *Arch Biochem Biophys.* **345**(1): 32-8.

Hu, X., O'Donnell, R., Srivastava, S. K., Xia, H., Zimniak, P., Nanduri, B., Bleicher, R. J., Awasthi, S., Awasthi, Y. C., Ji, X. and Singh, S. V. (1997b). Active site architecture of polymorphic forms of human glutathione S-transferase P1-1 accounts for their enantioselectivity and disparate activity in the glutathione conjugation of 7beta,8alpha-dihydroxy-9alpha,10alpha-ox y-7,8,9,10-tetrahydrobenzo(a)pyrene. *Biochem Biophys Res Commun.* **235**(2): 424-8.

Hubatsch, I., Ridderstrom, M. and Mannervik, B. (1998). Human glutathione transferase A4-4: an alpha class enzyme with high catalytic efficiency in the conjugation of 4-hydroxynonenal and other genotoxic products of lipid peroxidation. *Biochem J.* **330 (Pt 1)**: 175-9.

Hunter, E. S., 3rd, Rogers, E. H., Schmid, J. E. and Richard, A. (1996). Comparative effects of haloacetic acids in whole embryo culture. *Teratology.* **54**(2): 57-64.

Hussey, A. J., Kerr, L. A., Cronshaw, A. D., Harrison, D. J. and Hayes, J. D. (1991). Variation in the expression of Mu-class glutathione S-transferase isoenzymes from human skeletal muscle. Evidence for the existence of heterodimers. *Biochem J.* **273(Pt 2)**: 323-32.

Hussey, A. J. and Hayes, J. D. (1992). Characterization of a human class-Theta glutathione S-transferase with activity towards 1-menaphthyl sulphate. *Biochem J.* **286 (Pt 3)**: 929-35.

Iida, A., Saito, S., Sekine, A., Kitamoto, T., Kitamura, Y., Mishima, C., Osawa, S., Kondo, K., Harigae, S. and Nakamura, Y. (2001). Catalog of 434 single-nucleotide polymorphisms (SNPs) in genes of the alcohol dehydrogenase, glutathione S-transferase, and nicotinamide adenine dinucleotide, reduced (NADH) ubiquinone oxidoreductase families. *J Hum Genet.* **46**(7): 385-407.

Inskip, A., Elexperu-Camiruaga, J., Buxton, N., Dias, P. S., MacIntosh, J., Campbell, D., Jones, P. W., Yengi, L., Talbot, J. A., Strange, R. C. and et al. (1995). Identification of polymorphism at the glutathione S-transferase, GSTM3 locus: evidence for linkage with GSTM1*A. *Biochem J.* **312 (Pt 3)**: 713-6.

Irizarry, K., Kustanovich, V., Li, C., Brown, N., Nelson, S., Wong, W. and Lee, C. J. (2000). Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. *Nat Genet.* **26**(2): 233-6.

Ishigaki, S., Abramovitz, M. and Listowsky, I. (1989). Glutathione-S-transferases are major cytosolic thyroid hormone binding proteins. *Arch Biochem Biophys.* **273**(2): 265-72.

Islam, M. Q., Platz, A., Szpirer, J., Szpirer, C., Levan, G. and Mannervik, B. (1989). Chromosomal localization of human glutathione transferase genes of classes alpha, mu and pi. *Hum Genet.* **82**(4): 338-42.

Jahnke, V., Matthias, C., Fryer, A. and Strange, R. (1996). Glutathione S-transferase and cytochrome-P-450 polymorphism as risk factors for squamous cell carcinoma of the larynx. *Am J Surg.* **172**(6): 671-3.

Jakobsson, P. J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A. and Persson, B. (1999). Common structural features of MAPEG -- a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Sci.* **8**(3): 689-92.

Jakoby, W. B. (1978). The glutathione S-transferases: a group of multifunctional detoxification proteins. *Adv Enzymol Relat Areas Mol Biol.* **46**: 383-414.

Janicek, M. F., Haseltine, W. A. and Henner, W. D. (1985). Malondialdehyde precursors in gamma-irradiated DNA, deoxynucleotides and deoxynucleosides. *Nucleic Acids Res.* **13**(24): 9011-29.

Ji, X., Zhang, P., Armstrong, R. N. and Gilliland, G. L. (1992). The three-dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-A resolution. *Biochemistry.* **31**(42): 10169-84.

Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N. and Gilliland, G. L. (1995). Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods. *Biochemistry.* **34**(16): 5317-28.

Ji, X., von Rosenvinge, E. C., Johnson, W. W., Armstrong, R. N. and Gilliland, G. L. (1996). Location of a potential transport binding site in a sigma class glutathione transferase by x-ray crystallography. *Proc Natl Acad Sci U S A.* **93**(16): 8208-13.

Ji, X., Tordova, M., O'Donnell, R., Parsons, J. F., Hayden, J. B., Gilliland, G. L. and Zimniak, P. (1997). Structure and function of the xenobiotic substrate-binding site and location of a potential non-substrate-binding site in a class pi glutathione S-transferase. *Biochemistry.* **36**(32): 9690-702.

Johansson, A. S. and Mannervik, B. (2001a). Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. *J Biol Chem.* **276**(35): 33061-5.

Johansson, A. S. and Mannervik, B. (2001b). Interindividual variability of glutathione transferase expression. *In*: Interindividual variability in human drug metabolism. Pacifici, G. M. and Pelkonen, O. London, Taylor and Francis: Chapter 15.

Johansson, A.S. and Mannervik, B. (2002). Active-site residues governing high steroid isomerase activity in human glutathione transferase A3-3. *J Biol Chem.* **277**(19): 16648-54.

Johns, L. E. and Houlston, R. S. (2000). Glutathione S-transferase mu1 (GSTM1) status and bladder cancer risk: a meta-analysis. *Mutagenesis.* **15**(5): 399-404.

Johnson, W. W., Ueng, Y. F., Widersten, M., Mannervik, B., Hayes, J. D., Sherratt, P. J., Ketterer, B. and Guengerich, F. P. (1997). Conjugation of highly reactive aflatoxin B1 exo-8,9-epoxide catalyzed by rat and human glutathione transferases: estimation of kinetic parameters. *Biochemistry.* **36**(11): 3056-60.

Jourenkova-Mironova, N., Voho, A., Bouchardy, C., Wikman, H., Dayer, P., Benhamou, S. and Hirvonen, A. (1999a). Glutathione S-transferase GSTM1, GSTM3, GSTP1 and GSTT1 genotypes and the risk of smoking-related oral and pharyngeal cancers. *Int J Cancer.* **81**(1): 44-8.

Jourenkova-Mironova, N., Voho, A., Bouchardy, C., Wikman, H., Dayer, P., Benhamou, S. and Hirvonen, A. (1999b). Glutathione S-transferase GSTM3 and GSTP1 genotypes and larynx cancer risk. *Cancer Epidemiol Biomarkers Prev.* **8**(2): 185-8.

Jowsey, I. R., Thomson, A. M., Flanagan, J. U., Murdock, P. R., Moore, G. B., Meyer, D. J., Murphy, G. J., Smith, S. A. and Hayes, J. D. (2001). Mammalian class Sigma glutathione S-transferases: catalytic properties and tissue-specific expression of human and rat GSH-dependent prostaglandin D2 synthases. *Biochem J.* **359**(Pt 3): 507-16.

Juronen, E., Tasa, G., Uuskula, M., Pooga, M. and Mikelsaar, A. V. (1996). Purification, characterization and tissue distribution of human class theta glutathione S-transferase T1-1. *Biochem Mol Biol Int.* **39**(1): 21-9.

Kamisaka, K., Listowsky, I., Gatmaitan, Z. and Arias, I. M. (1975). Interactions of bilirubin and other ligands with ligandin. *Biochemistry.* **14**(10): 2175-80.

Kanaoka, Y., Ago, H., Inagaki, E., Nanayama, T., Miyano, M., Kikuno, R., Fujii, Y., Eguchi, N., Toh, H., Urade, Y. and Hayaishi, O. (1997). Cloning and crystal structure of hematopoietic prostaglandin D synthase. *Cell.* **90**(6): 1085-95.

Kanaoka, Y., Fujimori, K., Kikuno, R., Sakaguchi, Y., Urade, Y. and Hayaishi, O. (2000). Structure and chromosomal localization of human and mouse genes for hematopoietic prostaglandin D synthase. Conservation of the ancestral genomic structure of sigma-class glutathione S-transferase. *Eur J Biochem.* **267**(11): 3315-22.

Kangsadalampai, S., Chelvanayagam, G., Baker, R. T., Yenchitsomanus, P., Pung-amritt, P., Mahasandana, C. and Board, P. G. (1998). A novel Asn344 deletion in the core domain of coagulation factor XIII A subunit: its effects on protein structure and function. *Blood.* **92**(2): 481-7.

Kano, T., Sakai, M. and Muramatsu, M. (1987). Structure and expression of a human class pi glutathione S-transferase messenger RNA. *Cancer Res.* **47**(21): 5626-30.

Kato, T., Fukushima, M., Kurozumi, S. and Noyori, R. (1986). Antitumor activity of delta 7-prostaglandin A1 and delta 12-prostaglandin J2 in vitro and in vivo. *Cancer Res.* **46**(7): 3538-42.

Kawajiri, K., Nakachi, K., Imai, K., Yoshii, A., Shinoda, N. and Watanabe, J. (1990). Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. *FEBS Lett.* **263**(1): 131-3.

Keen, J., Lester, D., Inglehearn, C., Curtis, A. and Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.* **7**(1): 5.

Kelley, M. K., Engqvist-Goldstein, A., Montali, J. A., Wheatley, J. B., Schmidt, D. E., Jr. and Kauvar, L. M. (1994). Variability of glutathione S-transferase isoenzyme patterns in matched normal and cancer human breast tissue. *Biochem J.* **304 (Pt 3)**: 843-8.

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M. and Tsui, L. C. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science.* **245**(4922): 1073-80.

Ketterer, B. and Christodoulides, L. G. (1994). Enzymology of cytosolic glutathione S-transferases. *Adv Pharmacol.* **27**: 37-69.

Ketterer, B., Ross-Mansell, P. and Whitehead, J. K. (1967). The isolation of carcinogen-binding protein from livers of rats given 4-dimethylaminoazobenzene. *Biochem J.* **103**(2): 316-24.

Kihara, M., Noda, K. and Okamoto, N. (1993). Increased risk of lung cancer in Japanese smokers with class mu glutathione S-transferase gene deficiency. *Cancer Lett.* **71**(1-3): 151-5.

Kim, I. K., Lee, J. H., Sohn, H. W., Kim, H. S. and Kim, S. H. (1993). Prostaglandin A2 and delta 12-prostaglandin J2 induce apoptosis in L1210 cells. *FEBS Lett.* **321**(2-3): 209-14.

Kitahara, A., Satoh, K., Nishimura, K., Ishikawa, T., Ruike, K., Sato, K., Tsuda, H. and Ito, N. (1984). Changes in molecular forms of rat hepatic glutathione S-transferase during chemical hepatocarcinogenesis. *Cancer Res.* **44**(6): 2698-703.

Klaasen, C. D. (1996). Heavy Metals and Heavy-Metal Agonists. *In*: Goodman and Gilman's The Pharmacological Basis of Therapeutics. Hardman, J. G., Limbird, L. E., Molinoff, P. B., Ruddon, R. W. and Gilman, A. G., McGraw-Hill: 1649-72.

Kliewer, S. A., Lenhard, J. M., Willson, T. M., Patel, I., Morris, D. C. and Lehmann, J. M. (1995). A prostaglandin J2 metabolite binds peroxisome proliferator-activated receptor gamma and promotes adipocyte differentiation. *Cell.* **83**(5): 813-9.

Klone, A., Hussnatter, R. and Sies, H. (1992). Cloning, sequencing and characterization of the human alpha glutathione S-transferase gene corresponding to the cDNA clone pGTH2. *Biochem J.* **285 (Pt 3)**: 925-8.

Kodym, R., Calkins, P. and Story, M. (1999). The cloning and characterization of a new stress response protein. A mammalian member of a family of theta class glutathione s-transferase-like proteins. *J Biol Chem.* **274**(8): 5131-7.

Kono, M., Sen, A. C. and Chakrabarti, B. (1990). Thermodynamics of thermal and athermal denaturation of gamma-crystallins: changes in conformational stability upon glutathione reaction. *Biochemistry.* **29**(2): 464-70.

Kool, M., de Haas, M., Scheffer, G. L., Scheper, R. J., van Eijk, M. J., Juijn, J. A., Baas, F. and Borst, P. (1997). Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5, homologues of the multidrug resistance-associated protein gene (MRP1), in human cancer cell lines. *Cancer Res.* **57**(16): 3537-47.

Kool, M., van der Linden, M., de Haas, M., Baas, F. and Borst, P. (1999). Expression of human MRP6, a homologue of the multidrug resistance protein gene MRP1, in tissues and cancer cells. *Cancer Res.* **59**(1): 175-82.

Krawczak, M. and Cooper, D. N. (1991). Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum Genet.* **86**(5): 425-41.

Krawczak, M., Reiss, J. and Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet.* **90**(1-2): 41-54.

Krawczak, M., Ball, E. V., Fenton, I., Stenson, P. D., Abeysinghe, S., Thomas, N. and Cooper, D. N. (2000). Human gene mutation database-a biomedical information and research resource. *Hum Mutat.* **15**(1): 45-51.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet.* **22**(2): 139-44.

Kruglyak, L. and Nickerson, D. A. (2001). Variation is the spice of life. *Nat Genet.* **27**(3): 234-6.

Kumar, S. and Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature.* **392**(6679): 917-20.

Kwok, P. Y., Deng, Q., Zakeri, H., Taylor, S. L. and Nickerson, D. A. (1996). Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics.* **31**(1): 123-6.

Lai, E. (2001). Application of SNP technologies in medicine: lessons learned and future challenges. *Genome Res.* **11**(6): 927-9.

Laisney, V., Nguyen Van, C., Gross, M. S. and Frezal, J. (1984). Human genes for glutathione S-transferases. *Hum Genet.* **68**(3): 221-7.

Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988). A ligase-mediated gene detection technique. *Science.* **241**(4869): 1077-80.

Landegren, U., Nilsson, M. and Kwok, P. Y. (1998). Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res.* **8**(8): 769-76.

Lantum, H. B., Liebler, D. C., Board, P. G. and Anders, M. W. (2002). Alkylation and inactivation of human glutathione transferase zeta (hGSTZ1-1) by maleylacetone and fumarylacetone. *Chem Res Toxicol.* **15**(5): 707-16.

Latsoudis, H., Agapito, C., Hu, M. T. M., Chaudhuri, K. R. and Collie, r. D. A. (2000). Identification of a variable number of tandem repeats, two single nucleotide polymorphisms and a microsatellite repeat in the GSTA4 gene. Movement disorders. Barcelona, Spain. Abstract.

Layton, D. W., Bogen, K. T., Knize, M. G., Hatch, F. T., Johnson, V. M. and Felton, J. S. (1995). Cancer risk of heterocyclic amines in cooked foods: an analysis and implications for research. *Carcinogenesis.* **16**(1): 39-52.

Le Trong, I., Stenkamp, R. E., Ibarra, C., Atkins, W. M. and Adman, E. T. (2002). 1.3-A resolution structure of human glutathione S-transferase with S-hexyl glutathione bound reveals possible extended ligandin binding site. *Proteins.* **48**(4): 618-27.

Lee, S. G., Hong, S., Yoon, Y., Yang, I. and Song, K. (2001). Characterization of publicly available SNPs in the Korean population. *Hum Mutat.* **17**(4): 281-4.

Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M. B. (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics.* **33**(1): 151-2.

Lewis, A. D., Hickson, I. D., Robson, C. N., Harris, A. L., Hayes, J. D., Griffiths, S. A., Manson, M. M., Hall, A. E., Moss, J. E. and Wolf, C. R. (1988). Amplification and increased expression of alpha class glutathione S-transferase-encoding genes associated with resistance to nitrogen mustards. *Proc Natl Acad Sci U S A.* **85**(22): 8511-5.

Lewis, A. D., Forrester, L. M., Hayes, J. D., Wareing, C. J., Carmichael, J., Harris, A. L., Mooghen, M. and Wolf, C. R. (1989). Glutathione S-transferase isoenzymes in human tumours and tumour derived cell lines. *Br J Cancer.* **60**(3): 327-31.

Li, W. H. and Sadler, L. A. (1991). Low nucleotide diversity in man. *Genetics.* **129**(2): 513-23.

Lien, S., Larsson, A. K. and Mannervik, B. (2002). The polymorphic human glutathione transferase T1-1, the most efficient glutathione transferase in the denitrosation and inactivation of the anticancer drug 1,3-bis(2-chloroethyl)-1-nitrosourea. *Biochem Pharmacol.* **63**(2): 191-7.

Lin, D., Meyer, D. J., Ketterer, B., Lang, N. P. and Kadlubar, F. F. (1994). Effects of human and rat glutathione S-transferases on the covalent DNA binding of the N-acetoxy derivatives of heterocyclic amine carcinogens in vitro: a possible mechanism of organ specificity in their carcinogenesis. *Cancer Res.* **54**(18): 4920-6.

Lin, H. J., Probst-Hensch, N. M., Louie, A. D., Kau, I. H., Witte, J. S., Ingles, S. A., Frankl, H. D., Lee, E. R. and Haile, R. W. (1998). Glutathione transferase null genotype, broccoli, and lower prevalence of colorectal adenomas. *Cancer Epidemiol Biomarkers Prev.* **7**(8): 647-52.

Lindblad, B., Lindstedt, S. and Steen, G. (1977). On the enzymic defects in hereditary tyrosinemia. *Proc Natl Acad Sci U S A.* **74**(10): 4641-5.

Linder, R. E., Klinefelter, G. R., Strader, L. F., Narotsky, M. G., Suarez, J. D., Roberts, N. L. and Perreault, S. D. (1995). Dibromoacetic acid affects reproductive competence and sperm quality in the male rat. *Fundam Appl Toxicol.* **28**(1): 9-17.

Linder, R. E., Klinefelter, G. R., Strader, L. F., Suarez, J. D. and Roberts, N. L. (1997). Spermatotoxicity of dichloroacetic acid. *Reprod Toxicol.* **11**(5): 681-8.

Listowsky, I., Abramovitz, M., Homma, H. and Niitsu, Y. (1988). Intracellular binding and transport of hormones and xenobiotics by glutathione-S-transferases. *Drug Metab Rev.* **19**(3-4): 305-18.

Litwack, G., Ketterer, B. and Arias, I. M. (1971). Ligandin: a hepatic protein which binds steroids, bilirubin, carcinogens and a number of exogenous organic anions. *Nature.* **234**(5330): 466-7.

Liu, S., Stoesz, S. P. and Pickett, C. B. (1998). Identification of a novel human glutathione S-transferase using bioinformatics. *Arch Biochem Biophys.* **352**(2): 306-13.

Livak, K. J., Marmaro, J. and Todd, J. A. (1995). Towards fully automated genome-wide polymorphism screening. *Nat Genet.* **9**(4): 341-2.

Lorper, M., Schulz, W. A., Morel, F., Warskulat, U. and Sies, H. (1996). Positive and negative regulatory regions in promoters of human glutathione transferase alpha genes. *Biol Chem Hoppe Seyler.* **377**(1): 39-46.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951). Protein measurement with the Folin-Phenol reagents. *J. Biol. Chem.* **193**: 265-75.

Mannervik, B. and Danielson, U. H. (1988). Glutathione transferases--structure and catalytic activity. *CRC Crit Rev Biochem.* **23**(3): 283-337.

Mannervik, B. and Widersten, M. (1995). Human Glutathione Transferases: Classification, tissue distribution, structure and functional properties. *In*: Advances in drug metabolism in man. Pacifici, G. M. and Fracchia, G. N. Luxembourg, Luxembourg: European Commission: 409-59.

Mannervik, B., Guthenberg, C., Jakobson, I. and Warholm, M. (1978). Glutathione conjugation: reaction mechanism of glutathione S-transferase A. *In*: Conjugation Reactions in Drug Biotransformation. Aitio, A. Amsterdam, Elsevier/North Holland Biomedical Press: 101-10.

Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jornvall, H. (1985). Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties. *Proc Natl Acad Sci U S A.* **82**(21): 7202-6.

Mariot, P., Prevarskaya, N., Roudbaraki, M. M., Le Bourhis, X., Van Coppenolle, F., Vanoverberghe, K. and Skryma, R. (2000). Evidence of functional ryanodine receptor involved in apoptosis of prostate cancer (LNCaP) cells. *Prostate.* **43**(3): 205-14.

Mark, R. J., Lovell, M. A., Markesbery, W. R., Uchida, K. and Mattson, M. P. (1997). A role for 4-hydroxynonenal, an aldehydic product of lipid peroxidation, in disruption of ion homeostasis and neuronal death induced by amyloid beta-peptide. *J Neurochem.* **68**(1): 255-64.

Marsh, S., Kwok, P. and McLeod, H.L. (2002). SNP databases and pharmacogenetics: great start, but a long way to go. *Hum Mutat.* **20**(3): 174-9.

Marshall, E. (1997). 'Playing chicken' over gene markers. *Science.* **278**(5346): 2046-8.

Marshall, E. (1999). Drug firms to create public database of genetic mutations. *Science.* **284**(5413): 406-7.

Marth, G. T., Korf, I., Yandell, M. D., Yeh, R. T., Gu, Z., Zakeri, H., Stitziel, N. O., Hillier, L., Kwok, P. Y. and Gish, W. R. (1999). A general approach to single-nucleotide polymorphism discovery. *Nat Genet.* **23**(4): 452-6.

Marth, G., Yeh, R., Minton, M., Donaldson, R., Li, Q., Duan, S., Davenport, R., Miller, R. D. and Kwok, P. Y. (2001). Single-nucleotide polymorphisms in the public domain: how useful are they? *Nat Genet.* **27**(4): 371-2.

Masimirembwa, C. M., Dandara, C., Sommers, D. K., Snyman, J. R. and Hasler, J. A. (1998). Genetic polymorphism of cytochrome P4501A1, microsomal epoxide hydrolase, and glutathione S-transferases M1 and T1 in Zimbabweans and Venda of southern Africa. *Pharmacogenetics.* **8**(1): 83-5.

Matthias, C., Bockmuhl, U., Jahnke, V., Jones, P. W., Hayes, J. D., Alldersea, J., Gilford, J., Bailey, L., Bath, J., Worrall, S. F., Hand, P., Fryer, A. A. and Strange, R. C. (1998). Polymorphism in cytochrome P450 CYP2D6, CYP1A1, CYP2E1 and glutathione S-transferase, GSTM1, GSTM3, GSTT1 and susceptibility to tobacco-related cancers: studies in upper aerodigestive tract cancers. *Pharmacogenetics.* **8**(2): 91-100.

McLellan, R. A., Oscarson, M., Alexandrie, A. K., Seidegard, J., Evans, D. A., Rannug, A. and Ingelman-Sundberg, M. (1997). Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. *Mol Pharmacol.* **52**(6): 958-65.

McTigue, M. A., Williams, D. R. and Tainer, J. A. (1995). Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel. *J Mol Biol.* **246**(1): 21-7.

Meyer, D. J. and Ketterer, B. (1982). 5 alpha,6 alpha-Epoxy-cholestan-3 beta-ol (cholesterol alpha-oxide): A specific substrate for rat liver glutathione transferase B. *FEBS Lett.* **150**(2): 499-502.

Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991). Theta, a new class of glutathione transferases purified from rat and man. *Biochem J.* **274 (Pt 2)**: 409-14.

Meyer, D. J., Gilmore, K. S., Harris, J. M., Hartley, J. A. and Ketterer, B. (1992). Chlorambucil-monoglutathionyl conjugate is sequestered by human alpha class glutathione S-transferases. *Br J Cancer.* **66**(3): 433-8.

Miller, R. T., Christoffels, A. G., Gopalakrishnan, C., Burke, J., Ptitsyn, A. A., Broveak, T. R. and Hide, W. A. (1999). A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. *Genome Res.* **9**(11): 1143-55.

Minami, T., Uda, R., Horiguchi, S., Ito, S., Hyodo, M. and Hayaishi, O. (1994). Allodynia evoked by intrathecal administration of prostaglandin E2 to conscious mice. *Pain.* **57**(2): 217-23.

Morel, F., Fardel, O., Meyer, D. J., Langouet, S., Gilmore, K. S., Meunier, B., Tu, C. P., Kensler, T. W., Ketterer, B. and Guillouzo, A. (1993). Preferential increase of glutathione S-transferase class alpha transcripts in cultured human hepatocytes by phenobarbital, 3-methylcholanthrene, and dithiolethiones. *Cancer Res.* **53**(2): 231-4.

Morel, F., Rauch, C. and Guillouzo, A. (2001). Identification of a cDNA encoding a human omega class glutathione transferase (hGSTO2) and characterization of its gene structure. *Drug Metab Rev.* **33**(Suppl 1): 1-274.

Morel, F., Rauch, C., Coles, B., Ferrec, E. L. and Guillouzo, A. (2002). The human glutathione transferase alpha locus: genomic organization of the gene cluster and functional characterization of the genetic polymorphism in the hGSTA1 promoter. *Pharmacogenetics.* **12**(4): 277-86.

Morgenstern, R., Guthenberg, C. and Depierre, J. W. (1982). Microsomal glutathione S-transferase. Purification, initial characterization and demonstration that it is not identical to the cytosolic glutathione S-transferases A, B and C. *Eur J Biochem.* **128**(1): 243-8.

Morrow, C. S., Cowan, K. H. and Goldsmith, M. E. (1989). Structure of the human genomic glutathione S-transferase-pi gene. *Gene.* **75**(1): 3-11.

Moscow, J. A., Townsend, A. J., Goldsmith, M. E., Whang-Peng, J., Vickers, P. J., Poisson, R., Legault-Poisson, S., Myers, C. E. and Cowan, K. H. (1988). Isolation of the human anionic glutathione S-transferase cDNA and the relation of its gene expression to estrogen-receptor content in primary breast cancer. *Proc Natl Acad Sci U S A.* **85**(17): 6518-22.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**(2): 459-72.

Mountain, J. L., Lin, A. A., Bowcock, A. M. and Cavalli-Sforza, L. L. (1992). Evolution of modern humans: evidence from nuclear DNA polymorphisms. *Philos Trans R Soc Lond B Biol Sci.* **337**(1280): 159-65.

Mukanganyama, S., Masimirembwa, C. M., Naik, Y. S. and Hasler, J. A. (1997). Phenotyping of the glutathione S-transferase M1 polymorphism in Zimbabweans and the effects of chloroquine on blood glutathione S-transferases M1 and A. *Clin Chim Acta.* **265**(2): 145-55.

Mulder, T. P., Peters, W. H., Wobbes, T., Witteman, B. J. and Jansen, J. B. (1997). Measurement of glutathione S-transferase P1-1 in plasma: pitfalls and significance of screening and follow-up of patients with gastrointestinal carcinoma. *Cancer.* **80**(5): 873-80.

Mulder, T. P., Court, D. A. and Peters, W. H. (1999). Variability of glutathione S-transferase alpha in human liver and plasma. *Clin Chem.* **45**(3): 355-9.

Nachman, M. W., Bauer, V. L., Crowell, S. L. and Aquadro, C. F. (1998). DNA variability and recombination rates at X-linked loci in humans. *Genetics.* **150**(3): 1133-41.

Nakagawa, K., Saijo, N., Tsuchida, S., Sakai, M., Tsunokawa, Y., Yokota, J., Muramatsu, M., Sato, K., Terada, M. and Tew, K. D. (1990). Glutathione-S-transferase pi as a determinant of drug resistance in transfectant cell lines. *J Biol Chem.* **265**(8): 4296-301.

Nakajima, T., Elovaara, E., Anttila, S., Hirvonen, A., Camus, A. M., Hayes, J. D., Ketterer, B. and Vainio, H. (1995). Expression and polymorphism of glutathione S-transferase in human lungs: risk factors in smoking-related lung cancer. *Carcinogenesis.* **16**(4): 707-11.

Nei, M. and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* **76**(10): 5269-73.

Nei, M. and Roychoudhury, A. K. (1993). Evolutionary relationships of human populations on a global scale. *Mol Biol Evol.* **10**(5): 927-43.

Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C. and Markham, A. F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**(7): 2503-16.

Nickerson, D. A., Tobe, V. O. and Taylor, S. L. (1997). PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* **25**(14): 2745-51.

Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. and Sing, C. F. (1998). DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet.* **19**(3): 233-40.

Norppa, H. (1997). Cytogenetic markers of susceptibility: influence of polymorphic carcinogen-metabolizing enzymes. *Environ Health Perspect.* **105 Suppl 4**: 829-35.

Oakley, A. J., Lo Bello, M., Nuccetelli, M., Mazzetti, A. P. and Parker, M. W. (1999). The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site). *J Mol Biol.* **291**(4): 913-26.

Ogorochi, T., Narumiya, S., Mizuno, N., Yamashita, K., Miyazaki, H. and Hayaishi, O. (1984). Regional distribution of prostaglandins D2, E2, and F2 alpha and related enzymes in postmortem human brain. *J Neurochem.* **43**(1): 71-82.

Ogorochi, T., Ujihara, M. and Narumiya, S. (1987). Purification and properties of prostaglandin H-E isomerase from the cytosol of human brain: identification as anionic forms of glutathione S-transferase. *J Neurochem.* **48**(3): 900-9.

Ohnishi, K., Yoshida, H., Shigeno, K., Nakamura, S., Fujisawa, S., Naito, K., Shinjo, K., Fujita, Y., Matsui, H., Takeshita, A., Sugiyama, S., Satoh, H., Terada, H. and Ohno, R. (2000). Prolongation of the QT interval and ventricular tachycardia in patients treated with arsenic trioxide for acute promyelocytic leukemia. *Ann Intern Med.* **133**(11): 881-5.

Ohno, K. and Hirata, M. (1993). Characterization of the transport system of prostaglandin A2 in L-1210 murine leukemia cells. *Biochem Pharmacol.* **46**(4): 661-70.

Okuda, H., Miwa, K., Nojima, H. and Watabe, T. (1986). Inactivation of the carcinogen, 5-hydroxymethylchrysene, by glutathione conjugation via a sulphate ester in hepatic cytosol. *Biochem Pharmacol.* **35**(24): 4573-6.

Onoe, H., Watanabe, Y., Ono, K., Koyama, Y. and Hayaishi, O. (1992). Prostaglandin E2 exerts an awaking effect in the posterior hypothalamus at a site distinct from that mediating its febrile action in the anterior hypothalamus. *J Neurosci.* **12**(7): 2715-25.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A.* **86**(8): 2766-70.

Otsu, K., Willard, H. F., Khanna, V. K., Zorzato, F., Green, N. M. and MacLennan, D. H. (1990). Molecular cloning of cDNA encoding the Ca2+ release channel (ryanodine receptor) of rabbit cardiac muscle sarcoplasmic reticulum. *J Biol Chem.* **265**(23): 13472-83.

Ozawa, N. and Guengerich, F. P. (1983). Evidence for formation of an S-[2-(N7-guanyl)ethyl]glutathione adduct in glutathione-mediated binding of the carcinogen 1,2-dibromoethane to DNA. *Proc Natl Acad Sci U S A.* **80**(17): 5266-70.

Pacifici, G. M., Warholm, M., Guthenberg, C., Mannervik, B. and Rane, A. (1986). Organ distribution of glutathione transferase isoenzymes in the human fetus: differences between liver and extrahepatic tissues. *Biochem Pharmacol.* **35**(9): 1616-9.

Palinski, W., Rosenfeld, M. E., Yla-Herttuala, S., Gurtner, G. C., Socher, S. S., Butler, S. W., Parthasarathy, S., Carew, T. E., Steinberg, D. and Witztum, J. L. (1989). Low density lipoprotein undergoes oxidative modification in vivo. *Proc Natl Acad Sci U S A.* **86**(4): 1372-6.

Pan, Z., Damron, D., Nieminen, A. L., Bhat, M. B. and Ma, J. (2000). Depletion of intracellular Ca2+ by caffeine and ryanodine induces apoptosis of chinese hamster ovary cells transfected with ryanodine receptor. *J Biol Chem.* **275**(26): 19978-84.

Park, J. B. and Levine, M. (1996). Purification, cloning and expression of dehydroascorbic acid-reducing activity from human neutrophils: identification as glutaredoxin. *Biochem J.* **315 (Pt 3)**: 931-8.

Park, J. Y., Muscat, J. E., Ren, Q., Schantz, S. P., Harwick, R. D., Stern, J. C., Pike, V., Richie, J. P., Jr. and Lazarus, P. (1997). CYP1A1 and GSTM1 polymorphisms and oral cancer risk. *Cancer Epidemiol Biomarkers Prev.* **6**(10): 791-7.

Park, L. Y., Muscat, J. E., Kaur, T., Schantz, S. P., Stern, J. C., Richie, J. P., Jr. and Lazarus, P. (2000). Comparison of GSTM polymorphisms and risk for oral cancer between African-Americans and Caucasians. *Pharmacogenetics.* **10**(2): 123-31.

Parker, J. and Ankel, H. (1992). Formation of a prostaglandin A2-glutathione conjugate in L1210 mouse leukemia cells. *Biochem Pharmacol.* **43**(5): 1053-60.

Patskovsky, Y. V., Huang, M. Q., Takayama, T., Listowsky, I. and Pearson, W. R. (1999a). Distinctive structure of the human GSTM3 gene-inverted orientation relative to the mu class glutathione transferase gene cluster. *Arch Biochem Biophys.* **361**(1): 85-93.

Patskovsky, Y. V., Patskovska, L. N. and Listowsky, I. (1999b). Functions of His107 in the catalytic mechanism of human glutathione S-transferase hGSTM1a-1a. *Biochemistry.* **38**(4): 1193-202.

Patskovsky, Y. V., Patskovska, L. N. and Listowsky, I. (1999c). An asparagine-phenylalanine substitution accounts for catalytic differences between hGSTM3-3 and other human class mu glutathione S-transferases. *Biochemistry.* **38**(49): 16187-94.

Patskovsky, Y. V., Patskovska, L. N. and Listowsky, I. (2000). The enhanced affinity for thiolate anion and activation of enzyme-bound glutathione is governed by an arginine residue of human Mu class glutathione S-transferases. *J Biol Chem.* **275**(5): 3296-304.

Pearson, W. R., Vorachek, W. R., Xu, S. J., Berger, R., Hart, I., Vannais, D. and Patterson, D. (1993). Identification of class-mu glutathione transferase genes GSTM1-GSTM5 on human chromosome 1p13. *Am J Hum Genet.* **53**(1): 220-33.

Pelkonen, O., Vahakangas, K. and Nebert, D. W. (1980). Binding of polycyclic aromatic hydrocarbons to DNA: comparison with mutagenesis and tumorigenesis. *J Toxicol Environ Health.* **6**(5-6): 1009-20.

Pemble, S. E. and Taylor, J. B. (1992). An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences. *Biochem J.* **287 (Pt 3)**: 957-63.

Pemble, S., Schroeder, K. R., Spencer, S. R., Meyer, D. J., Hallier, E., Bolt, H. M., Ketterer, B. and Taylor, J. B. (1994). Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem J.* **300 (Pt 1)**: 271-6.

Pemble, S. E., Wardle, A. F. and Taylor, J. B. (1996). Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. *Biochem J.* **319 (Pt 3)**: 749-54.

Peter, H., Deutschmann, S., Muelle, A., Gansewendt, B., Bolt, M. and Hallier, E. (1989). Different affinity of erythrocyte glutathione-S-transferase to methyl chloride in humans. *Arch Toxicol Suppl.* **13**: 128-32.

Peterson, G. L. (1977). A simplification of the protein assay method of Lowry et al. which is more generally applicable. *Anal Biochem.* **83**(2): 346-56.

Petrick, J. S., Ayala-Fierro, F., Cullen, W. R., Carter, D. E. and Vasken Aposhian, H. (2000). Monomethylarsonous acid (MMA(III)) is more toxic than arsenite in Chang human hepatocytes. *Toxicol Appl Pharmacol.* **163**(2): 203-7.

Pettigrew, N. E. and Colman, R. F. (2001). Heterodimers of glutathione S-transferase can form between isoenzyme classes pi and mu. *Arch Biochem Biophys.* **396**(2): 225-30.

Piccolomini, R., Di Ilio, C., Aceto, A., Allocati, N., Faraone, A., Cellini, L., Ravagnan, G. and Federici, G. (1989). Glutathione transferase in bacteria: subunit composition and antigenic characterization. *J Gen Microbiol.* **135 (Pt 11)**: 3119-25.

Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A. and Boyce-Jacino, M. (1999). Mining SNPs from EST databases. *Genome Res.* **9**(2): 167-74.

Ploemen, J. H., Van Schanke, A., Van Ommen, B. and Van Bladeren, P. J. (1994). Reversible conjugation of ethacrynic acid with glutathione and human glutathione S-transferase P1-1. *Cancer Res.* **54**(4): 915-9.

Polekhina, G., Board, P. G., Blackburn, A. C. and Parker, M. W. (2001). Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity. *Biochemistry.* **40**(6): 1567-76.

Prade, L., Huber, R., Manoharan, T. H., Fahl, W. E. and Reuter, W. (1997). Structures of class pi glutathione S-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. *Structure.* **5**(10): 1287-95.

Quinones, L., Berthou, F., Varela, N., Simon, B., Gil, L. and Lucas, D. (1999). Ethnic susceptibility to lung cancer: differences in CYP2E1, CYP1A1 and GSTM1 genetic polymorphisms between French Caucasian and Chilean populations. *Cancer Lett.* **141**(1-2): 167-71.

Radabaugh, T. R., Sampayo-Reyes, A., Zakharyan, R. A. and Aposhian, H. V. (2002). Arsenate Reductase II. Purine Nucleoside Phosphorylase in the Presence of Dihydrolipoic Acid Is a Route for Reduction of Arsenate to Arsenite in Mammalian Systems. *Chem Res Toxicol.* **15**(5): 692-8.

Raghunathan, S., Chandross, R. J., Kretsinger, R. H., Allison, T. J., Penington, C. J. and Rule, G. S. (1994). Crystal structure of human class mu glutathione transferase GSTM2-2. Effects of lattice packing on conformational heterogeneity. *J Mol Biol.* **238**(5): 815-32.

Raijmakers, M. T., Bruggeman, S. W., Steegers, E. A. and Peters, W. H. (2002). Distribution of components of the glutathione detoxification system across the human placenta after uncomplicated vaginal deliveries. *Placenta.* **23**(6): 490-6.

Ramachandran, S., Lear, J. T., Ramsay, H., Smith, A. G., Bowers, B., Hutchinson, P. E., Jones, P. W., Fryer, A. A. and Strange, R. C. (1999). Presentation with multiple cutaneous basal cell carcinomas: association of glutathione S-transferase and cytochrome P450 genotypes with clinical phenotype. *Cancer Epidemiol Biomarkers Prev.* **8**(1): 61-7.

Ramachandran, S., Fryer, A. A., Smith, A. G., Lear, J. T., Bowers, B., Hartland, A. J., Whiteside, J. R., Jones, P. W. and Strange, R. C. (2001). Basal cell carcinomas: association of allelic variants with a high-risk subgroup of patients with the multiple presentation phenotype. *Pharmacogenetics.* **11**(3): 247-54.

Ramensky, V., Bork, P. and Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**(17): 3894-900.

Rebbeck, T. R. (1997). Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.* **6**(9): 733-43.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J. and Lancet, D. (1997). GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.* **13**(4): 163.

Reginato, M. J., Krakow, S. L., Bailey, S. T. and Lazar, M. A. (1998). Prostaglandins promote and block adipogenesis through opposing effects on peroxisome proliferator-activated receptor gamma. *J Biol Chem.* **273**(4): 1855-8.

Reinemer, P., Dirr, H. W., Ladenstein, R., Huber, R., Lo Bello, M., Federici, G. and Parker, M. W. (1992). Three-dimensional structure of class pi glutathione S-transferase from human placenta in complex with S-hexylglutathione at 2.8 A resolution. *J Mol Biol.* **227**(1): 214-26.

Retief, J. D., Lynch, K. R. and Pearson, W. R. (1999). Panning for genes--A visual strategy for identifying novel gene orthologs and paralogs. *Genome Res.* **9**(4): 373-82.

Rhoads, D. M., Zarlengo, R. P. and Tu, C. P. (1987). The basic glutathione S-transferases from human livers are products of separate genes. *Biochem Biophys Res Commun.* **145**(1): 474-81.

Ricci, G., Caccuri, A. M., Lo Bello, M., Pastore, A., Piemonte, F. and Federici, G. (1994). Colorimetric and fluorometric assays of glutathione transferase based on 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole. *Anal Biochem.* **218**(2): 463-5.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature.* **405**(6788): 847-56.

Riva, A. and Kohane, I. S. (2001). A web-based tool to retrieve human genome polymorphisms from public databases. *Proc AMIA Symp.* 558-62.

Robertson, I. G., Guthenberg, C., Mannervik, B. and Jernstrom, B. (1986). Differences in stereoselectivity and catalytic efficiency of three human glutathione transferases in the conjugation of glutathione with 7 beta,8 alpha-dihydroxy-9 alpha,10 alpha-oxy-7,8,9,10-tetrahydrobenzo(a)pyrene. *Cancer Res.* **46**(5): 2220-4.

Rohrdanz, E., Nguyen, T. and Pickett, C. B. (1992). Isolation and characterization of the human glutathione S-transferase A2 subunit gene. *Arch Biochem Biophys.* **298**(2): 747-52.

Rommens, J. M., Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N. and et al. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science.* **245**(4922): 1059-65.

Roscoe, H. E. (1862). On the alleged practice of arsenic-eating in Styria. *Mem Lit Philos Soc Manchester, London Ser 3.* **1**: 208-21.

Ross, V. L. and Board, P. G. (1993). Molecular cloning and heterologous expression of an alternatively spliced human Mu class glutathione S-transferase transcript. *Biochem J.* **294 (Pt 2)**: 373-80.

Ross, V. L., Board, P. G. and Webb, G. C. (1993). Chromosomal mapping of the human Mu class glutathione S-transferases to 1p13. *Genomics.* **18**(1): 87-91.

Rossjohn, J., Board, P. G., Parker, M. W. and Wilce, M. C. (1996). A structurally derived consensus pattern for theta class glutathione transferases. *Protein Eng.* **9**(4): 327-32.

Rossjohn, J., McKinstry, W. J., Oakley, A. J., Verger, D., Flanagan, J., Chelvanayagam, G., Tan, K. L., Board, P. G. and Parker, M. W. (1998). Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site. *Structure.* **6**(3): 309-22.

Rowe, J. D., Nieves, E. and Listowsky, I. (1997). Subunit diversity and tissue distribution of human glutathione S-transferases: interpretations based on electrospray ionization-MS and peptide sequence-specific antisera. *Biochem J.* **325 (Pt 2)**: 481-6.

Rozen, F., Nguyen, T. and Pickett, C. B. (1992). Isolation and characterization of a human glutathione S-transferase Ha1 subunit gene. *Arch Biochem Biophys.* **292**(2): 589-93.

Ryberg, D., Hewer, A., Phillips, D. H. and Haugen, A. (1994). Different susceptibility to smoking-induced DNA damage among male and female lung cancer patients. *Cancer Res.* **54**(22): 5801-3.

Ryberg, D., Skaug, V., Hewer, A., Phillips, D. H., Harries, L. W., Wolf, C. R., Ogreid, D., Ulvik, A., Vu, P. and Haugen, A. (1997). Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. *Carcinogenesis.* **18**(7): 1285-9.

Salagovic, J., Kalina, I., Habalova, V., Hrivnak, M., Valansky, L. and Biros, E. (1999). The role of human glutathione S-transferases M1 and T1 in individual susceptibility to bladder cancer. *Physiol Res.* **48**(6): 465-71.

Sampayo-Reyes, A., Zakharyan, R. A., Healy, S. M. and Aposhian, H. V. (2000). Monomethylarsonic acid reductase and monomethylarsonous acid in hamster tissue. *Chem Res Toxicol.* **13**(11): 1181-6.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* **74**(12): 5463-7.

Sato, K. (1989). Glutathione transferases as markers of preneoplasia and neoplasia. *Adv Cancer Res.* **52**: 205-55.

Sayed, Y., Hornby, J. A., Lopez, M. and Dirr, H. (2002). Thermodynamics of the ligandin function of human class Alpha glutathione transferase A1-1: energetics of organic anion ligand binding. *Biochem J.* **363**(Pt 2): 341-6.

Schuler, G. D. (1997). Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J Mol Med.* **75**(10): 694-8.

Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T. J. and et al. (1996). A gene map of the human genome. *Science.* **274**(5287): 540-6.

Segura-Aguilar, J., Baez, S., Widersten, M., Welch, C. J. and Mannervik, B. (1997). Human class Mu glutathione transferases, in particular isoenzyme M2-2, catalyze detoxication of the dopamine metabolite aminochrome. *J Biol Chem.* **272**(9): 5727-31.

Sei, Y., Gallagher, K. L. and Basile, A. S. (1999). Skeletal muscle type ryanodine receptor is involved in calcium signaling in human B lymphocytes. *J Biol Chem.* **274**(9): 5995-6002.

Seidegard, J. and Ekstrom, G. (1997). The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics. *Environ Health Perspect.* **105 Suppl 4**: 791-9.

Seidegard, J., Pero, R. W., Miller, D. G. and Beattie, E. J. (1986). A glutathione transferase in human leukocytes as a marker for the susceptibility to lung cancer. *Carcinogenesis.* **7**(5): 751-3.

Seidegard, J., Vorachek, W. R., Pero, R. W. and Pearson, W. R. (1988). Hereditary differences in the expression of the human glutathione transferase active on trans-stilbene oxide are due to a gene deletion. *Proc Natl Acad Sci U S A.* **85**(19): 7293-7.

Seidegard, J., Pero, R. W., Markowitz, M. M., Roush, G., Miller, D. G. and Beattie, E. J. (1990). Isoenzyme(s) of glutathione transferase (class Mu) as a marker for the susceptibility to lung cancer: a follow up study. *Carcinogenesis.* **11**(1): 33-6.

Seres, T., Ravichandran, V., Moriguchi, T., Rokutan, K., Thomas, J. A. and Johnston, R. B., Jr. (1996). Protein S-thiolation and dethiolation during the respiratory burst in human monocytes. A reversible post-translational modification with potential for buffering the effects of oxidant stress. *J Immunol.* **156**(5): 1973-80.

Shea, T. C., Kelley, S. L. and Henner, W. D. (1988). Identification of an anionic form of glutathione transferase present in many human tumors and human tumor cell lines. *Cancer Res.* **48**(3): 527-33.

Sheehan, D., Meade, G., Foley, V. M. and Dowd, C. A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J.* **360**(Pt 1): 1-16.

Sheffield, V. C., Cox, D. R., Lerman, L. S. and Myers, R. M. (1989). Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci U S A.* **86**(1): 232-6.

Shen, Z. X., Chen, G. Q., Ni, J. H., Li, X. S., Xiong, S. M., Qiu, Q. Y., Zhu, J., Tang, W., Sun, G. L., Yang, K. Q., Chen, Y., Zhou, L., Fang, Z. W., Wang, Y. T., Ma, J., Zhang, P., Zhang, T. D., Chen, S. J., Chen, Z. and Wang, Z. Y. (1997). Use of arsenic trioxide (As2O3) in the treatment of acute promyelocytic leukemia (APL): II. Clinical efficacy and pharmacokinetics in relapsed patients. *Blood.* **89**(9): 3354-60.

Sherratt, P. J., Pulford, D. J., Harrison, D. J., Green, T. and Hayes, J. D. (1997). Evidence that human class Theta glutathione S-transferase T1-1 can catalyse the activation of dichloromethane, a liver and lung carcinogen in the mouse. Comparison of the tissue distribution of GST T1-1 with that of classes Alpha, Mu and Pi GST in human. *Biochem J.* **326 (Pt 3)**: 837-46.

Sherratt, P. J., Williams, S., Foster, J., Kernohan, N., Green, T. and Hayes, J. D. (2002). Direct comparison of the nature of mouse and human GST T1-1 and the implications on dichloromethane carcinogenicity. *Toxicol Appl Pharmacol.* **179**(2): 89-97.

Sherry, S. T., Ward, M. and Sirotkin, K. (1999). dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **9**(8): 677-9.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**(1): 308-11.

Shi, Y., Seto, E., Chang, L. S. and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell.* **67**(2): 377-88.

Shields, P. G., Bowman, E. D., Harrington, A. M., Doan, V. T. and Weston, A. (1993). Polycyclic aromatic hydrocarbon-DNA adducts in human lung and cancer susceptibility genes. *Cancer Res.* **53**(15): 3486-92.

Shuber, A. P., Michalowsky, L. A., Nass, G. S., Skoletsky, J., Hire, L. M., Kotsopoulos, S. K., Phipps, M. F., Barberio, D. M. and Klinger, K. W. (1997). High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes. *Hum Mol Genet.* **6**(3): 337-47.

Silverberg, E. and Lubera, J. (1987). Cancer statistics, 1987. *CA Cancer J Clin.* **37**(1): 2-19.

Simons, P. C. and Jagt, D. L. (1980). Bilirubin binding to human liver ligandin (glutathione S-transferase). *J Biol Chem.* **255**(10): 4740-4.

Singhal, S. S., Saxena, M., Ahmad, H., Awasthi, S., Haque, A. K. and Awasthi, Y. C. (1992a). Glutathione S-transferases of human lung: characterization and evaluation of the protective role of the alpha-class isozymes against lipid peroxidation. *Arch Biochem Biophys.* **299**(2): 232-41.

Singhal, S. S., Saxena, M., Awasthi, S., Ahmad, H., Sharma, R. and Awasthi, Y. C. (1992b). Gender related differences in the expression and characteristics of glutathione S-transferases of human colon. *Biochim Biophys Acta.* **1171**(1): 19-26.

Singhal, S. S., Saxena, M., Awasthi, S., Mukhtar, H., Zaidi, S. I., Ahmad, H. and Awasthi, Y. C. (1993). Glutathione S-transferases of human skin: qualitative and quantitative differences in men and women. *Biochim Biophys Acta.* **1163**(3): 266-72.

Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G. and et al. (1993). Structure determination and refinement of human alpha class glutathione transferase A1-1, and a comparison with the Mu and Pi class enzymes. *J Mol Biol.* **232**(1): 192-212.

Smedley, P. L. and Kinniburgh, D. G. (2001). Source and behaviour of arsenic in natural waters. *In*: United Nations Synthesis Report on Arsenic in Drinking Water**:** Chapter 1.

Soares, M. B., Bonaldo, M. F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci U S A.* **91**(20): 9228-32.

Soignet, S. L., Maslak, P., Wang, Z. G., Jhanwar, S., Calleja, E., Dardashti, L. J., Corso, D., DeBlasio, A., Gabrilove, J., Scheinberg, D. A., Pandolfi, P. P. and Warrell, R. P., Jr. (1998). Complete remission after treatment of acute promyelocytic leukemia with arsenic trioxide. *N Engl J Med.* **339**(19): 1341-8.

Soignet, S. L., Frankel, S. R., Douer, D., Tallman, M. S., Kantarjian, H., Calleja, E., Stone, R. M., Kalaycio, M., Scheinberg, D. A., Steinherz, P., Sievers, E. L., Coutre, S., Dahlberg, S., Ellison, R. and Warrell, R. P., Jr. (2001). United States multicenter study of arsenic trioxide in relapsed acute promyelocytic leukemia. *J Clin Oncol.* **19**(18): 3852-60.

Stacpoole, P. W. (1989). The pharmacology of dichloroacetate. *Metabolism.* **38**(11): 1124-44.

Stadtman, E. R. (1992). Protein oxidation and aging. *Science.* **257**(5074): 1220-4.

Stanulla, M., Schrappe, M., Brechlin, A. M., Zimmermann, M. and Welte, K. (2000). Polymorphisms within glutathione S-transferase genes (GSTM1, GSTT1, GSTP1) and risk of relapse in childhood B-cell precursor acute lymphoblastic leukemia: a case-control study. *Blood.* **95**(4): 1222-8.

Stenberg, G., Board, P. G., Carlberg, I. and Mannervik, B. (1991a). Effects of directed mutagenesis on conserved arginine residues in a human Class Alpha glutathione transferase. *Biochem J.* **274 (Pt 2)**: 549-55.

Stenberg, G., Board, P. G. and Mannervik, B. (1991b). Mutation of an evolutionarily conserved tyrosine residue in the active site of a human class Alpha glutathione transferase. *FEBS Lett.* **293**(1-2): 153-5.

Stevens, J. M., Hornby, J. A., Armstrong, R. N. and Dirr, H. W. (1998). Class sigma glutathione transferase unfolds via a dimeric and a monomeric intermediate: impact of subunit interface on conformational stability in the superfamily. *Biochemistry.* **37**(44): 15534-41.

Stevens, J. M., Armstrong, R. N. and Dirr, H. W. (2000). Electrostatic interactions affecting the active site of class sigma glutathione S-transferase. *Biochem J.* **347 Pt 1**: 193-7.

Stockman, P. K., Beckett, G. J. and Hayes, J. D. (1985). Identification of a basic hybrid glutathione S-transferase from human liver. Glutathione S-transferase delta is composed of two distinct subunits (B1 and B2). *Biochem J.* **227**(2): 457-65.

Stockman, P. K., McLellan, L. I. and Hayes, J. D. (1987). Characterization of the basic glutathione S-transferase B1 and B2 subunits from human liver. *Biochem J.* **244**(1): 55-61.

Stoesser, G., Baker, W., van den Broek, A., Camon, E., Garcia-Pastor, M., Kanz, C., Kulikova, T., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Redaschi, N., Stoehr, P., Tuli, M. A., Tzouvara, K. and Vaughan, R. (2002). The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **30**(1): 21-6.

Strange, R. C., Faulder, C. G., Davis, B. A., Hume, R., Brown, J. A., Cotton, W. and Hopkinson, D. A. (1984). The human glutathione S-transferases: studies on the tissue distribution and genetic variation of the GST1, GST2 and GST3 isozymes. *Ann Hum Genet.* **48 (Pt 1)**: 11-20.

Strange, R. C., Davis, B. A., Faulder, C. G., Cotton, W., Bain, A. D., Hopkinson, D. A. and Hume, R. (1985). The human glutathione S-transferases: developmental aspects of the GST1, GST2, and GST3 loci. *Biochem Genet.* **23**(11-12): 1011-28.

Strange, R. C., Howie, A. F., Hume, R., Matharoo, B., Bell, J., Hiley, C., Jones, P. and Beckett, G. J. (1989). The development expression of alpha-, mu- and pi-class glutathione S-transferases in human liver. *Biochim Biophys Acta.* **993**(2-3): 186-90.

Strange, R. C., Matharoo, B., Faulder, G. C., Jones, P., Cotton, W., Elder, J. B. and Deakin, M. (1991). The human glutathione S-transferases: a case-control study of the incidence of the GST1 0 phenotype in patients with adenocarcinoma. *Carcinogenesis.* **12**(1): 25-8.

Strange, R. C., Jones, P. W. and Fryer, A. A. (2000). Glutathione S-transferase: genetics and role in toxicology. *Toxicol Lett.* **112-113**: 357-63.

Strange, R. C., Spiteri, M. A., Ramachandran, S. and Fryer, A. A. (2001). Glutathione-S-transferase family of enzymes. *Mutat Res.* **482**(1-2): 21-6.

Stroombergen, M. C. and Waring, R. H. (1999). Determination of glutathione S-transferase mu and theta polymorphisms in neurological disease. *Hum Exp Toxicol.* **18**(3): 141-5.

Sundberg, K., Johansson, A. S., Stenberg, G., Widersten, M., Seidel, A., Mannervik, B. and Jernstrom, B. (1998a). Differences in the catalytic efficiencies of allelic variants of glutathione transferase P1-1 towards carcinogenic diol epoxides of polycyclic aromatic hydrocarbons. *Carcinogenesis.* **19**(3): 433-6.

Sundberg, K., Seidel, A., Mannervik, B. and Jernstrom, B. (1998b). Detoxication of carcinogenic fjord-region diol epoxides of polycyclic aromatic hydrocarbons by glutathione transferase P1-1 variants and glutathione. *FEBS Lett.* **438**(3): 206-10.

Sutton, G. G., White, O., Adams, M. D. and Kerlavage, A. R. (1995). TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Science and Technology.* **1**(1): 9-19.

Suzuki, T., Coggan, M., Shaw, D. C. and Board, P. G. (1987). Electrophoretic and immunological analysis of human glutathione S-transferase isozymes. *Ann Hum Genet.* **51 (Pt 2)**: 95-106.

Suzuki, T., Johnston, P. N. and Board, P. G. (1993). Structure and organization of the human alpha class glutathione S-transferase genes and related pseudogenes. *Genomics.* **18**(3): 680-6.

Syvanen, A. C. (1999). From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat.* **13**(1): 1-10.

Taioli, E., Crofts, F., Trachman, J., Demopoulos, R., Toniolo, P. and Garte, S. J. (1995a). A specific African-American CYP1A1 polymorphism is associated with adenocarcinoma of the lung. *Cancer Res.* **55**(3): 472-3.

Taioli, E., Trachman, J., Chen, X., Toniolo, P. and Garte, S. J. (1995b). A CYP1A1 restriction fragment length polymorphism is associated with breast cancer in African-American women. *Cancer Res.* **55**(17): 3757-8.

Takahashi, Y., Campbell, E. A., Hirata, Y., Takayama, T. and Listowsky, I. (1993). A basis for differentiating among the multiple human Mu-glutathione S-transferases and molecular cloning of brain GSTM5. *J Biol Chem.* **268**(12): 8893-8.

Takeshima, H., Nishimura, S., Matsumoto, T., Ishida, H., Kangawa, K., Minamino, N., Matsuo, H., Ueda, M., Hanaoka, M., Hirose, T. and et al. (1989). Primary structure and expression from complementary DNA of skeletal muscle ryanodine receptor. *Nature.* **339**(6224): 439-45.

Tan, K. L. and Board, P. G. (1996). Purification and characterization of a recombinant human Theta-class glutathione transferase (GSTT2-2). *Biochem J.* **315 (Pt 3)**: 727-32.

Tan, K. L., Webb, G. C., Baker, R. T. and Board, P. G. (1995). Molecular cloning of a cDNA and chromosomal localization of a human theta-class glutathione S-transferase gene (GSTT2) to chromosome 22. *Genomics.* **25**(2): 381-7.

Tang, D. L., Rundle, A., Warburton, D., Santella, R. M., Tsai, W. Y., Chiamprasert, S., Hsu, Y. Z. and Perera, F. P. (1998). Associations between both genetic and environmental biomarkers and lung cancer: evidence of a greater risk of lung cancer in women smokers. *Carcinogenesis.* **19**(11): 1949-53.

Taniguchi, K., Wada, M., Kohno, K., Nakamura, T., Kawabe, T., Kawakami, M., Kagotani, K., Okumura, K., Akiyama, S. and Kuwano, M. (1996). A human canalicular multispecific organic anion transporter (cMOAT) gene is overexpressed in cisplatin-resistant human cancer cell lines with decreased drug accumulation. *Cancer Res.* **56**(18): 4124-9.

Tateno, Y., Imanishi, T., Miyazaki, S., Fukami-Kobayashi, K., Saitou, N., Sugawara, H. and Gojobori, T. (2002). DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic Acids Res.* **30**(1): 27-30.

Taylor, J. B., Oliver, J., Sherrington, R. and Pemble, S. E. (1991). Structure of human glutathione S-transferase class Mu genes. *Biochem J.* **274 (Pt 2)**: 587-93.

Tefre, T., Ryberg, D., Haugen, A., Nebert, D. W., Skaug, V., Brogger, A. and Borresen, A. L. (1991). Human CYP1A1 (cytochrome P(1)450) gene: lack of association between the Msp I restriction fragment length polymorphism and incidence of lung cancer in a Norwegian population. *Pharmacogenetics.* **1**(1): 20-5.

Tew, K. D. (1994). Glutathione-associated enzymes in anticancer drug resistance. *Cancer Res.* **54**(16): 4313-20.

Thier, R., Pemble, S. E., Kramer, H., Taylor, J. B., Guengerich, F. P. and Ketterer, B. (1996). Human glutathione S-transferase T1-1 enhances mutagenicity of 1,2-dibromoethane, dibromomethane and 1,2,3,4-diepoxybutane in Salmonella typhimurium. *Carcinogenesis.* **17**(1): 163-6.

Tiainen, P. and Karhi, K.K. (1994). Ultrasensitive time-resolved immunofluorometric assay of glutathione transferase alpha in serum. *Clin Chem.* **40**(2): 184-9.

To-Figueras, J., Gene, M., Gomez-Catalan, J., Pique, E., Borrego, N., Marfany, G., Gonzalez Duarte, R. and Corbella, J. (2000). Polymorphism of glutathione S-transferase M3: interaction with glutathione S-transferase M1 and lung cancer susceptibility. *Biomarkers.* **5**(1): 73-80.

Tong, Z., Board, P. G. and Anders, M. W. (1998a). Glutathione transferase zeta-catalyzed biotransformation of dichloroacetic acid and other alpha-haloacids. *Chem Res Toxicol.* **11**(11): 1332-8.

Tong, Z., Board, P. G. and Anders, M. W. (1998b). Glutathione transferase zeta catalyses the oxygenation of the carcinogen dichloroacetic acid to glyoxylic acid. *Biochem J.* **331 (Pt 2)**: 371-4.

Tseng, W. P. (1977). Effects and dose--response relationships of skin cancer and blackfoot disease with arsenic. *Environ Health Perspect.* **19**: 109-19.

Tsuchida, S. and Sato, K. (1992). Glutathione transferases and cancer. *Crit Rev Biochem Mol Biol.* **27**(4-5): 337-84.

Tsuchida, S., Maki, T. and Sato, K. (1990). Purification and characterization of glutathione transferases with an activity toward nitroglycerin from human aorta and heart. Multiplicity of the human class Mu forms. *J Biol Chem.* **265**(13): 7150-7.

Tu, C. P. and Qian, B. (1986). Human liver glutathione S-transferases: complete primary sequence of an Ha subunit cDNA. *Biochem Biophys Res Commun.* **141**(1): 229-37.

Tyagi, S., Bratu, D. P. and Kramer, F. R. (1998). Multicolor molecular beacons for allele discrimination. *Nat Biotechnol.* **16**(1): 49-53.

Tzeng, H. F., Blackburn, A. C., Board, P. G. and Anders, M. W. (2000). Polymorphism- and species-dependent inactivation of glutathione transferase zeta by dichloroacetate. *Chem Res Toxicol.* **13**(4): 231-6.

Uden, P. C. and Miller, J. W. (1983). Chlorinated acids and chloral in drinking water. *J Am Water Works Assoc.* **75**: 524-7.

Ueno, R., Narumiya, S., Ogorochi, T., Nakayama, T., Ishikawa, Y. and Hayaishi, O. (1982). Role of prostaglandin D2 in the hypothermia of rats caused by bacterial lipopolysaccharide. *Proc Natl Acad Sci U S A.* **79**(19): 6093-7.

Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. and Cavalli-Sforza, L. L. (1996). A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci U S A.* **93**(1): 196-200.

Urade, Y., Fujimoto, N., Ujihara, M. and Hayaishi, O. (1987). Biochemical and immunological characterization of rat spleen prostaglandin D synthetase. *J Biol Chem.* **262**(8): 3820-5.

Vahter, M., Concha, G., Nermell, B., Nilsson, R., Dulout, F. and Natarajan, A. T. (1995). A unique metabolism of inorganic arsenic in native Andean women. *Eur J Pharmacol.* **293**(4): 455-62.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E.,

Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001). The sequence of the human genome. *Science.* **291**(5507): 1304-51.

Vineis, P. (1994). Epidemiology of cancer from exposure to arylamines. *Environ Health Perspect.* **102 Suppl 6**: 7-10.

Vorachek, W. R., Pearson, W. R. and Rule, G. S. (1991). Cloning, expression, and characterization of a class-mu glutathione transferase from human muscle, the product of the GST4 locus. *Proc Natl Acad Sci U S A.* **88**(10): 4443-7.

Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W. and Povey, S. (2002). Guidelines for human gene nomenclature. *Genomics.* **79**(4): 464-70.

Wang, Z. and Moult, J. (2001). SNPs, protein structure, and disease. *Hum Mutat.* **17**(4): 263-70.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lander, E. S. and et al. (1998). Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. *Science.* **280**(5366): 1077-82.

Wang, A., C., W., Coles B., K.E., A., N.P., L. and F.F., K. (2000). Identification and characterisation of a novel polymorphic variant of hGSTA2-2. Proceedings of the American Association for Cancer Research. San Francisco, USA. Abstract.

Warholm, M., Guthenberg, C., Mannervik, B., von Bahr, C. and Glaumann, H. (1980). Identification of a new glutathione S-transferase in human liver. *Acta Chem Scand B.* **34**(8): 607-21.

Warholm, M., Guthenberg, C., Mannervik, B., Pacifici, G. M. and Rane, A. (1981). Glutathione S-transferases in human fetal liver. *Acta Chem Scand B.* **35**(3): 225-7.

Warholm, M., Guthenberg, C. and Mannervik, B. (1983). Molecular and catalytic properties of glutathione transferase mu from human liver: an enzyme efficiently conjugating epoxides. *Biochemistry.* **22**(15): 3610-7.

Washko, P. W., Wang, Y. and Levine, M. (1993). Ascorbic acid recycling in human neutrophils. *J Biol Chem.* **268**(21): 15531-5.

Watabe, T., Hiratsuka, A. and Ogura, K. (1986). Regioselective glutathione conjugation of the carcinogen, 7, 12-dihydroxymethylbenz[a]anthracene, via reactive 7-hydroxymethyl sulfate ester in rat liver cytosol. *Biochem Biophys Res Commun.* **134**(1): 100-5.

Webb, G., Vaska, V., Coggan, M. and Board, P. (1996). Chromosomal localization of the gene for the human theta class glutathione transferase (GSTT1). *Genomics.* **33**(1): 121-3.

Weiss, K. M. (1998). In search of human variation. *Genome Res.* **8**(7): 691-7.

Welfare, M., Monesola Adeokun, A., Bassendine, M. F. and Daly, A. K. (1999). Polymorphisms in GSTP1, GSTM1, and GSTT1 and susceptibility to colorectal cancer. *Cancer Epidemiol Biomarkers Prev.* **8**(4 Pt 1): 289-92.

Wempe, M. F., Anderson, W. B., Tzeng, H. F., Board, P. G. and Anders, M. W. (1999). Glutathione transferase zeta-catalyzed biotransformation of deuterated dihaloacetic acids. *Biochem Biophys Res Commun.* **261**(3): 779-83.

Westervelt, P., Brown, R. A., Adkins, D. R., Khoury, H., Curtin, P., Hurd, D., Luger, S. M., Ma, M. K., Ley, T. J. and DiPersio, J. F. (2001). Sudden death among patients with acute promyelocytic leukemia treated with arsenic trioxide. *Blood.* **98**(2): 266-71.

Whalen, R. and Boyer, T. D. (1998). Human glutathione S-transferases. *Semin Liver Dis.* **18**(4): 345-58.

Wheeler, J. B., Stourman, N. V., Armstrong, R. N. and Guengerich, F. P. (2001a). Conjugation of haloalkanes by bacterial and mammalian glutathione transferases: mono- and vicinal dihaloethanes. *Chem Res Toxicol.* **14**(8): 1107-17.

Wheeler, J. B., Stourman, N. V., Thier, R., Dommermuth, A., Vuilleumier, S., Rose, J. A., Armstrong, R. N. and Guengerich, F. P. (2001b). Conjugation of haloalkanes by bacterial and mammalian glutathione transferases: mono- and dihalomethanes. *Chem Res Toxicol.* **14**(8): 1118-27.

Whitbread, A. K., Tetlow, N., Eyre, H. J., Sutherland, G. R. and Board, P. G. Characterisation of the human Omega class glutathione transferase genes and associated polymorphisms. *Biochem J.* In Press.

Widersten, M., Pearson, W. R., Engstrom, A. and Mannervik, B. (1991). Heterologous expression of the allelic variant mu-class glutathione transferases mu and psi. *Biochem J.* **276 (Pt 2)**: 519-24.

Wilce, M. C. and Parker, M. W. (1994). Structure and function of glutathione S-transferases. *Biochim Biophys Acta.* **1205**(1): 1-18.

Wilce, M. C., Board, P. G., Feil, S. C. and Parker, M. W. (1995). Crystal structure of a theta-class glutathione transferase. *Embo J.* **14**(10): 2133-43.

Williamson, A. R. (1999). The Merck Gene Index project. *Drug Discov Today.* **4**(3): 115-22.

Wittenberger, T., Schaller, H. C. and Hellebrand, S. (2001). An expressed sequence tag (EST) data mining strategy succeeding in the discovery of new G-protein coupled receptors. *J Mol Biol.* **307**(3): 799-813.

Wolkoff, A. W., Goresky, C. A., Sellin, J., Gatmaitan, Z. and Arias, I. M. (1979). Role of ligandin in transfer of bilirubin from plasma into liver. *Am J Physiol.* **236**(6): E638-48.

Wong, L. C., Winston, J. M., Hong, C. B. and Plotnick, H. (1982). Carcinogenicity and toxicity of 1,2-dibromoethane in the rat. *Toxicol Appl Pharmacol.* **63**(2): 155-65.

Worrall, S. F., Corrigan, M., High, A., Starr, D., Matthias, C., Wolf, C. R., Jones, P. W., Hand, P., Gilford, J., Farrell, W. E., Hoban, P., Fryer, A. A. and Strange, R. C. (1998). Susceptibility and outcome in oral cancer: preliminary data showing an association with polymorphism in cytochrome P450 CYP2D6. *Pharmacogenetics.* **8**(5): 433-9.

Xia, C., Taylor, J. B., Spencer, S. R. and Ketterer, B. (1993). The human glutathione S-transferase P1-1 gene: modulation of expression by retinoic acid and insulin. *Biochem J.* **292 (Pt 3)**: 845-50.

Xu, L., Tripathy, A., Pasek, D. A. and Meissner, G. (1998a). Potential for pharmacology of ryanodine receptor/calcium release channels. *Ann N Y Acad Sci.* **853**: 130-48.

Xu, S., Wang, Y., Roe, B. and Pearson, W. R. (1998b). Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion. *J Biol Chem.* **273**(6): 3517-27.

Yamamura, S., Bartram, J., Csanady, M., Gorchev, H. G. and Redekopp, A. (2001). Drinking Water Guidelines and Standards. *In*: United Nations Synthesis Report on Arsenic in Drinking Water: Chapter 5.

Yengi, L., Inskip, A., Gilford, J., Alldersea, J., Bailey, L., Smith, A., Lear, J. T., Heagerty, A. H., Bowers, B., Hand, P., Hayes, J. D., Jones, P. W., Strange, R. C. and Fryer, A. A. (1996). Polymorphism at the glutathione S-transferase locus GSTM3: interactions with cytochrome P450 and glutathione S-transferase genotypes as risk factors for multiple cutaneous basal cell carcinoma. *Cancer Res.* **56**(9): 1974-7.

Yoritaka, A., Hattori, N., Uchida, K., Tanaka, M., Stadtman, E. R. and Mizuno, Y. (1996). Immunohistochemical detection of 4-hydroxynonenal protein adducts in Parkinson disease. *Proc Natl Acad Sci U S A.* **93**(7): 2696-701.

Youil, R., Kemper, B. and Cotton, R. G. (1996). Detection of 81 of 81 known mouse beta-globin promoter mutations with T4 endonuclease VII--the EMC method. *Genomics.* **32**(3): 431-5.

Zakharyan, R. A. and Aposhian, H. V. (1999). Enzymatic reduction of arsenic compounds in mammalian systems: the rate-limiting enzyme of rabbit liver arsenic biotransformation is MMA(V) reductase. *Chem Res Toxicol.* **12**(12): 1278-83.

Zakharyan, R. A., Sampayo-Reyes, A., Healy, S. M., Tsaprailis, G., Board, P. G., Liebler, D. C. and Aposhian, H. V. (2001). Human monomethylarsonic acid (MMA(V)) reductase is a member of the glutathione-S-transferase superfamily. *Chem Res Toxicol.* **14**(8): 1051-7.

Zhang, G., Zhang, S., Chen, W., Qiu, W., Wu, H., Wang, J., Luo, J., Gu, X. and Cotton, R. G. (2001). Go!Poly: A gene-oriented polymorphism database. *Hum Mutat.* **18**(5): 382-7.

Zhao, L., Alldersea, J., Fryer, A., Tighe, A., Ollier, B., Thomson, W., Jones, P. and Strange, R. (1994). Polymorphism at the glutathione S-transferase GSTM1 locus: a study of the frequencies of the GSTM1 A, B, A/B and null phenotypes in Nigerians. *Clin Chim Acta.* **225**(1): 85-8.

Zhao, T., Singhal, S. S., Piper, J. T., Cheng, J., Pandya, U., Clark-Wronski, J., Awasthi, S. and Awasthi, Y. C. (1999). The role of human glutathione S-transferases hGSTA1-1 and hGSTA2-2 in protection against oxidative stress. *Arch Biochem Biophys.* **367**(2): 216-24.

Zhong, S., Howie, A. F., Ketterer, B., Taylor, J., Hayes, J. D., Beckett, G. J., Wathen, C. G., Wolf, C. R. and Spurr, N. K. (1991). Glutathione S-transferase mu locus: use of genotyping and phenotyping assays to assess association with lung cancer susceptibility. *Carcinogenesis.* **12**(9): 1533-7.

Zhong, S., Spurr, N. K., Hayes, J. D. and Wolf, C. R. (1993a). Deduced amino acid sequence, gene structure and chromosomal location of a novel human class Mu glutathione S-transferase, GSTM4. *Biochem J.* **291 (Pt 1)**: 41-50.

Zhong, S., Wyllie, A. H., Barnes, D., Wolf, C. R. and Spurr, N. K. (1993b). Relationship between the GSTM1 genetic polymorphism and susceptibility to bladder, breast and colon cancer. *Carcinogenesis.* **14**(9): 1821-4.

Zietkiewicz, E., Yotova, V., Jarnik, M., Korab-Laskowska, M., Kidd, K. K., Modiano, D., Scozzari, R., Stoneking, M., Tishkoff, S., Batzer, M. and Labuda, D. (1997). Nuclear DNA diversity in worldwide distributed human populations. *Gene.* **205**(1-2): 161-71.

Zorzato, F., Fujii, J., Otsu, K., Phillips, M., Green, N. M., Lai, F. A., Meissner, G. and MacLennan, D. H. (1990). Molecular cloning of cDNA encoding human and rabbit forms of the Ca2+ release channel (ryanodine receptor) of skeletal muscle sarcoplasmic reticulum. *J Biol Chem.* **265**(4): 2244-56.

APPENDICES

Appendix 1. Common Reagents and their Sources

REAGENT	SOURCE
α^{33}P[dATP]	Amersham Pharmacia Biotech (UK)
Acetic Acid	BDH AnalaR, Merck Pty Ltd (Australia)
Accugel 29:1™: 40% (29:1) Acrylamide/Bisacrylamide	National Diagnostics (USA)
Agarose, Type II	BDH AnalaR, Merck Pty Ltd (Australia)
Alkaline Protease	Sigma-Aldrich (USA)
Ammonium Persulphate	BDH Laboratory Supplies (UK)
Ampicillin	Sigma-Aldrich (USA)
Δ^5-Androsten-3,17-dione	Steraloids Inc. (USA)
β-mercaptoethanol	BDH AnalaR, Merck Pty Ltd (Australia)
Bacto-agar	Bacto Laboratories Pty Ltd (Australia)
Black and White Film	Polaroid Corp. (USA)
Boric Acid	BioRad (USA)
Bovine Serum Albumin, Pierce Standard	PIERCE (USA)
Calcium Chloride	Mallinckrodt Inc. (USA)
Chloramphenicol	Sigma-Aldrich (USA)
1-chloro-2,4-dinitrobenzene	Fluka AG, Buchs SG (Switzerland)
Chloroform	APS Ajax Finechem (Australia)
7-chloro-4-nitrobenzo-2-oxa-1,3-diazole	Sigma-Aldrich (USA)
Coomassie Brilliant Blue	BDH AnalaR, Merck Pty Ltd (Australia)
Copper Sulphate	BDH AnalaR, Merck Pty Ltd (Australia)
Cumene Hydroperoxide	Aldrich (USA)
Deoxycholic Acid	Sigma-Aldrich (USA)
2'-deoxynucleoside 5'-triphosphate	Life Technologies (USA)
1,2-dichloro-nitrobenzene	Sigma-Aldrich (USA)
2-dithioerythriotol	Sigma-Aldrich (USA)
Ethacrynic Acid	Sigma-Aldrich (USA)

1,2-epoxy-3(-4-nitrophenoxy)propane	Sigma-Aldrich (USA)
Ethylenediaminetetraacetic acid	BDH AnalaR, Merck Pty Ltd (Australia)
Ethanol	BDH AnalaR, Merck Pty Ltd (Australia)
Ethidium Bromide	Sigma-Aldrich (USA)
Ficoll	Amersham Pharmacia Biotech (UK)
Folin & Ciocalteu phenol reagent	Sigma-Aldrich (USA)
Glucose	BDH AnalaR, Merck Pty Ltd (Australia)
Glutathione Agarose	Sigma-Aldrich (USA)
Glutathione, reduced	Sigma-Aldrich (USA)
Glutathione Reductase	Sigma-Aldrich (USA)
Glycerol	APS Ajax Finechem (Australia)
Glycine	APS Ajax Finechem (Australia)
Hydrochloric Acid	BDH AnalaR, Merck Pty Ltd (Australia)
Isopropylthio-β-D-galactosidase	Progen Industries Ltd (Australia)
λ DNA/*Hind*III	Promega Corp. (USA)
Manganese Chloride	APS Ajax Finechem (Australia)
Methanol	BDH AnalaR, Merck Pty Ltd (Australia)
p-nitrobenzyl chloride	Sigma-Aldrich (USA)
p-nitrophenylacetate	Sigma-Aldrich (USA)
One Kb DNA Ladder	Life Technologies (USA)
One Kb Plus DNA Ladder	Life Technologies (USA)
Phenol	Wako Pure Chemical Industries (Japan)
4-phenyl-but-3-en-2-one	Aldrich (USA)
Potassium Acetate	BDH AnalaR, Merck Pty Ltd (Australia)
Potassium Chloride	APS Ajax Finechem (Australia)
Protogel™: 37.5:1 acrylamide:bisacrylamide	National Diagnostics (USA)
pUC/*Hpa*II Ladder	GeneWorks (Australia)
Rainbow™ Marker	Amersham Life Science (UK)
Sequagel™: 19:1 acrylamide:bisacrylamide	National Diagnostics (USA)

Sequagel XR concentrate	National Diagnostics (USA)
Sodium Acetate	BDH AnalaR, Merck Pty Ltd (Australia)
Sodium Carbonate	BDH AnalaR, Merck Pty Ltd (Australia)
Sodium Chloride	BDH AnalaR, Merck Pty Ltd (Australia)
Sodium di-hydrogen orthophosphate	Fronine Pty Ltd (Australia)
Sodium Dodecyl Sulphate	BDH AnalaR, Merck Pty Ltd (Australia)
Sodium Hydroxide	Merck (Germany)
di-Sodium hydrogen orthophosphate	BDH AnalaR, Merck Pty Ltd (Australia)
Sodium-MOPS	Sigma-Aldrich (USA)
Sodium nicotinamide	Boehringer Mannheim (Germany)
Sodium Tartrate	Mallinckrodt Inc. (USA)
Sucrose	APS Ajax Finechem (Australia)
Taurine	Merck Schuchardt (Germany)
TEMED	Sigma-Aldrich (USA)
Tert-butyl hydroperoxide	Sigma-Aldrich (USA)
Trans-non-2-enal	Aldrich (USA)
Trichoroacetic acid	BDH AnalaR, Merck Pty Ltd (Australia)
Tris	Life Technologies (USA)
Tryptone	Bacto Laboratories Pty Ltd (Australia)
Urea	BDH AnalaR, Merck Pty Ltd (Australia)
Super RX Medical X Ray Film	Fuji, Hanimex Pty Ltd (Australia)
Yeast Extract	Bacto Laboratories Pty Ltd (Australia)

Appendix 2. Enzymes and their Sources

ENZYME	SOURCE
Restriction Endonucleases and 10x Buffers	Roche Diagnostics (Germany)
	New England Biolabs (U.S.A.)
	MBI Fermentas (Lithuania)
	Amersham Pharmacia Biotech (U.K.)
Pfu Turbo DNA Polymerase and 10x Buffer	Stratagene (U.S.A.)
Pfu DNA Polymerase and 10x Buffer	Stratagene (U.S.A.)
T4 DNA Ligase	Amersham Pharmacia Biotech (UK)
	Promega (U.S.A.)
Taq DNA Polymerase and 10x Buffer	ABGene (U.K.)
	Promega Corp. (U.S.A.)

Appendix 3. Buffers, Media and Solutions

Table A3.1 – Media

Media	**Components**
Luria Broth (LB)	1% tryptone
	0.5% yeast extract
	1% NaCl, pH7.4
LB plates	1.5% (w/v) agar
	LB
LB-Amp plates	1.5% (w/v) agar
	100 µg/ml Ampicillin
	LB
LB-Amp/Chlor plates	1.5% (w/v) agar
	100 µg/ml Ampicillin
	34 µg/mL Chloramphenicol
	LB

Table A3.2 – Buffers for Making Competent Cells

Buffers	Components
TFBI	30 mM KAc
	50 mM $MnCl_2.4H_2O$
	100 mM KCl
	10 mM $CaCl_2.2H_2O$
	15% glycerol
TFBII	10 mM Na-MOPS pH7
	75 mM $CaCl_2.2H_2O$
	10 mM KCl
	15% glycerol

Table A3.3 – Buffers and Solutions for DNA Work

Solution	**Components**
5x BRL Buffer	250 mM Tris-HCl pH7.6
	50 mM $MgCl_2$
	25% (w/v) PEG
	5 mM ATP
	5 mM DTT
10x Glycerol Tolerant Buffer	900 mM Tris base
	300 mM Taurine
	5 mM EDTA
Ficoll gel loading dye	0.25% bromophenol blue
	15% ficoll
	0.25% xylene cyanol FF
Sucrose gel loading dye	0.25% Bromophenol Blue
	40% (w/v) sucrose
50x TAE buffer	2 M Tris-Acetate
	50 mM EDTA, pH 8.0
1x TBE buffer	900 mM Tris-Borate
	20 mM EDTA, pH 8.0
TE buffer	10 mM Tris-HCl pH7.5 or 8.0
	1 mM EDTA

Table A3.4 – Buffers and Solutions used for Protein Work

Solution	**Component**
4x Resolving Gel Buffer	1.5 M Tris pH8.8
	0.4% SDS
4x Stacking Gel buffer	500 mM Tris pH 6.8
	0.4% SDS
10x Running Buffer	50 mM Tris base
	192 M Glycine
	1% SDS
3x SDS gel loading Buffer	187.5 mM Tris-HCl pH6.8
	6% SDS (w/v)
	30% glycerol (v/v)
	0.0075% bromophenol blue
	1 mM β-mercaptoethanol
Staining Solution	10 mL Coomassie® Blue R250
	50 mL destaining solution
Destaining Solution	10% acetic acid
	30% methanol

Appendix 4. Preliminary genotyping data for polymorphisms detected in the Zeta and Theta class GSTs

Table A4.1 – *GSTT1* and *GSTZ1* allele frequencies in three ethnic groups

Gene	Polymorphism	Population	n	Genotype			Allele Frequency	
GSTT1	p.V95L	Australian	25	G/G=22	G/T=3	T/T=0	p.V95=0.94	p.L95=0.06
GSTZ1	p.N133H	Australian	100	A/A=100	A/C=0	C/C=0	p.N133=1	p.H133=0
		Bantu	68	A/A=100	A/C=0	C/C=0	p.N133=1	p.H133=0
		Chinese	99	A/A=94	A/C=5	C/C=0	p.N133=0.98	p.H133=0.2